

## SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: RITA MITRA Examiner #: 7795 Date: 9/10/03  
An Unit: 1653 Phone Number 30 605-1211 Serial Number: 091737297  
Mail Box and Bldg/Room Location: 9801/CM1 Results Format Preferred (circle): PAPER DISK E-MAIL  
9803

If more than one search is submitted, please prioritize searches in order of need.

\*\*\*\*\*  
Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc. if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: "Processes and Organisms for the production of Anti-freeze proteins"

Inventors (please provide full names): MARK JOHN BERRY, ALLEN GRIFFITHS, PHILIP JOHN HILL, JOHANNA LAYBOURNE-PARRY, SARAH VICTORIA MILLS

Earliest Priority Filing Date: December 15, 1999

\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

I would request an expedited literature search for above case (Patent & Non-Patent both). Not sequence search

The search should encompass "anti-freeze proteins" (AFPs), having anti-freeze property. Search should be focussed on isoforms and derivatives of these proteins; food products comprising a protein showing anti-freeze property and also anti-freeze protein isolated from bacterial culture.

Keywords:

Antifreeze property, inhibition of recrystallisation. *Bacteria Marinomonas sp. and Pseudomonas sp.*

Elected Claims: — 4, 5, 6, 7, 11, 12

Point of Contact:

Toby Port

Technical Info. Specialist

CM1 6A04

703-308-3554

C. Chan  
Rush

## STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher: \_\_\_\_\_ NA Sequence (#) \_\_\_\_\_ STN 204

Citation Page: \_\_\_\_\_ AA Sequence (#) \_\_\_\_\_ Dialog \_\_\_\_\_

Tue May 13 16:15:50 2003

us-09-737-297-3.rag

Page 1

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK WPB - 12/15/1997 All are NO/EP Rdg

OW protein - protein search, using sw model  
Run on: May 13, 2003, 15:51:29 ; Search time 75 Seconds  
(without alignments)  
28.427 Million cell updates/sec

Title: US-09-737-297-3  
Perfect score: 79  
Sequence: 1 AEGSTDPVYONIQAC 16

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77	97.5	16	AA84488	N-terminal of anti
2	47	59.5	22	ABG12516	Novel human diago-
3	43	54.4	380	AB854010	Lactococcus lactis
4	42	53.2	231	AA44183	Streptomyces oliva
5	41	51.9	191	AA60738	Xylanase C of stre
6	41	51.9	191	AA60281	Modified xylanase
7	41	51.9	191	AA59675	S. lividans xylana
8	41	51.9	191	AA818447	Streptomyces livid
9	41	51.9	200	AA832263	Endo-xylanase from
10	41	51.9	240	AA848535	Streptomyces livid

11	41	51.9	359	22	ABG13264	Novel human diago
12	41	51.9	541	22	ABG08970	Novel human diago
13	41	51.9	1560	22	ABG05639	Novel human diago
14	41	51.9	2659	22	ABG68426	Drosophila melanog
15	40	50.6	313	21	AA71906	7G5 mouse rH2AF1
16	40	50.6	456	22	AAU35277	Enterococcus faeca
17	40	50.6	672	21	AA81257	Moloney murine leu
18	40	50.6	683	23	AAU74990	Moloney murine leu
19	40	50.6	683	23	AAU74991	Moloney murine leu
20	40	50.6	683	23	AAU74992	Moloney murine leu
21	40	50.6	683	23	AAU74994	Moloney murine leu
22	40	50.6	683	23	AAU74995	Moloney murine leu
23	40	50.6	683	23	AAU74997	Moloney murine leu
24	40	50.6	683	23	AAU74998	Moloney murine leu
25	40	50.6	683	23	AAU74999	Moloney murine leu
26	40	50.6	683	23	AAU75004	Moloney murine leu
27	40	50.6	683	23	AAU75008	Moloney murine leu
28	40	50.6	683	23	AAU75010	Moloney murine leu
29	40	50.6	683	23	AAU75012	Moloney murine leu
30	40	50.6	683	23	AAU75013	Moloney murine leu
31	40	50.6	683	23	AAU75014	Moloney murine leu
32	40	50.6	683	23	AAU75017	Moloney murine leu
33	40	50.6	683	23	AAU75020	Moloney murine leu
34	40	50.6	683	23	AAU75020	Moloney murine leu
35	40	50.6	683	23	AAU75020	Moloney murine leu
36	40	50.6	683	23	AAU75020	Moloney murine leu
37	39	49.4	15	18	AAW01788	Residual protease
38	39	49.4	313	11	AAW04905	Bacillus subtilis
39	39	49.4	313	18	AAW16358	Bacillus subtilis
40	39	49.4	313	12	AAW01791	Residual protease
41	39	49.4	313	22	AAW00017	Bacillus subtilis
42	39	49.4	379	23	ABP30295	Streptococcus poly
43	39	49.4	385	23	ABP37594	Streptococcus poly
44	39	49.4	432	22	ABG02271	Novel human diago
45	39	49.4	576	22	ABG2520	Drosophila melanog

## ALIGNMENTS

RESULT 1	AA84488	standard; peptide: 16 AA.
ID	AA84488	
XX	AA84488	
AC	05-SEP-2001 (first entry)	
XX		
DT		
DE	N-terminal of anti-freeze peptide of Marinomonas protea.	
XX		
KW	16S rRNA; anti-freeze protein; food additive; frozen vegetable;	
KW	frozen confectiometry.	
XX		
OS	Marinomonas protea.	
XX		
FX		
FT	Key	Location/Qualifiers
FT	Misc-difference	6 /label- Gly, Val
XX		
PN	WO200144275-A2.	
XX		
PD	21-JUN-2001.	
XX		
PF	05-DEC-2000; 2000WO-EP12396.	
XX		
PR	15-DEC-1999; 99GB-0029696.	
XX		
PA	(UNIT.) UNILEVER PLC.	
PA	(UNIT.) UNILEVER NV.	
PA	(HIND-) HINDUSTAN LEVER LTD.	
XX		
PI	Berry MJ, Griffiths A, Hill PJ, Laybourne-Parry J, Mills SV;	
XX		

DR WPI; 2001-398120/42.  
XX  
XX Preparing anti-freeze peptides useful in frozen food products, e.g.  
PT frozen vegetables and confectionery, by culturing bacteria from aqueous  
PT low-temperature environment and extracting anti-freeze proteins from  
PT culture  
XX  
XX Claim 7; Page 9; 59pp; English.  
XX  
CC The present sequence represents the N-terminal of an anti-freeze  
CC peptide (AFP) of Marinomonas protea. Marinomonas protea. AFPS can be  
CC isolated from M. protea, using the method of the invention. The  
CC specification describes a method for producing AFPS. The method  
CC comprises collecting one or more samples of bacteria from an aqueous  
CC low-temperature environment, culturing the bacteria and extracting  
CC proteins from the samples, testing the proteins for anti-freeze  
CC properties, selecting proteins having anti-freeze properties and  
CC producing the selected protein for use as an AFP food additive. The  
CC method is useful for producing AFPS which are incorporated in food  
CC products, such as frozen vegetables and frozen confectionery such as  
CC ice-cream. AFPS are useful in frozen food products, such as vegetables,  
CC sauces, soups, snacks, dairy products and frozen confectionery, which  
CC includes sorbet, water-ice, granites, frozen fruit purees and  
CC milk-containing frozen products such as ice-cream, frozen yogurt or  
CC custards, sherbet and ice-milk.  
XX  
XX Sequence 16 AA:  
SQ  
Query Match 97.5%; Score 77; DB 22; Length 16;  
Best Local Similarity 100.0%; Pred. No. 6,6e-08;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AEGSTXDYQNIQYAG 16  
DB 1 AEGSTXDYQNIQYAG 16  
RESULT 2  
ABG12516  
ID ABG12516 standard; Protein; 1164 AA.  
XX  
XX ABG12516;  
AC  
XX  
XX 18-FEB-2002 (first entry)  
DT  
XX  
XX Novel human diagnostic protein #12507.  
DE  
XX  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
CS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 30-MAR-2001; 2001WO-US06631.  
PF  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
DR  
XX  
XX N-PSDB; AAS76703.  
DR  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX

PS Claim 20; SEQ ID No 42875; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1164 AA:  
SQ  
Query Match 59.5%; Score 47; DB 22; Length 1164;  
Best Local Similarity 53.3%; Pred. No. 7.9;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
OY 2 EGSTXDYQNIQYAG 16  
DB 13 DGSFAEFYQRIKYLIG 27  
RESULT 3  
ABBS4010  
ID ABBS4010 standard; Protein; 380 AA.  
XX  
XX ABBS4010;  
AC  
XX  
XX 16-MAY-2002 (first entry)  
DT  
XX  
XX Lactococcus lactis protein glgc.  
DE  
XX  
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
KW  
XX  
XX Lactococcus lactis IL1403.  
CS  
XX  
XX FR2807446-A1.  
PN  
XX  
XX 12-OCT-2001.  
PD  
XX  
XX 11-APR-2000; 2000FR-0004630.  
PF  
XX  
XX 11-APR-2000; 2000FR-0004630.  
PR  
XX  
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.  
PA  
XX  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
PI  
XX  
XX WPI; 2002-043418/06.  
DR  
XX  
XX New nucleotide sequence useful in the identification of Lactococcus  
PT lactis and related species  
PT  
XX  
XX Claim 6; SEQ ID No 712; 2504pp; French.  
PS  
XX  
XX The present invention is related to a Lactococcus lactis nucleotide  
CC sequence (ABR90521) and related proteins (ABBS53100-ABBS56211). The  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
CC related species. The proteins of the invention are useful for the

CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO200177334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.  
 CC  
 SQ Sequence 380 AA;

Query Match 54.4%; Score 43; DB 23; Length 380;  
 Best Local Similarity 61.5%; Pred. No. 11;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 EGGTXDVYQNIQY 14  
 DB 100 EGGTXDVYQNIQY 112

## RESULT 4

ID AAY44183 standard; Protein; 231 AA.

AC AAY44183;

DT 01-FEB-2000 (first entry)

DE Streptomyces olivaceoviridis xylanase (Xyng) protein.

KW Xylanase; plasmid; expression; E.coli; xylO-oligosaccharide; xylan;  
 KW pulp-bleaching.

OS Streptococcus olivaceoviridis.

XX  
 FH Key

FT Peptide 1..40 Location/Qualifiers

FT Protein 41..999 /label= signal\_peptide

FT /label= mature\_protein  
 FT /note= "the mature protein and its coding sequence  
 are claimed"

PN JPI1266873-A.

PD 05-OCT-1999

PF 20-MAR-1998; 98JP-0090702.

PR 20-MAR-1998; 98JP-0090702.

PA (NORO) NORINSUISANSHO SHOKUHN SOGO.

PA (SEIB-) SEIBUTSUKEI TOKUTEI SANGYO GIJUTSU.

DR WPI: 1999-613780/53.

DR N-PSDB; AA228864.

PT A xylanase gene, contiguous with a vector and a transformant - used for  
 PT pulp-bleaching

PS Example 1; Page 5-6; 10pp; Japanese.

CC This sequence corresponds to the xyng xylanase Streptomyces  
 CC olivaceoviridis. The coding sequence can be inserted into the plasmid  
 CC pOE60 to generate plasmid FERM P-16713 for expression in e.g. E.coli.  
 CC The xylanase is useful for the preparation of xylO-oligosaccharide from  
 CC xylan and for pulp-bleaching.

SQ Sequence 231 AA;

Query Match 53.2%; Score 42; DB 20; Length 231;

Best Local Similarity 50.0%; Pred. No. 9.3;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYQNIQY 14

DB 149 SDGCTYDVYQITRY 162

RESULT 5  
 ID AAM60738 standard; Protein; 191 AA.  
 AC AAM60738;

DT 02-SEP-1998 (first entry)

DE Xylanase C of Streptomyces lividans.

KW Family 11 xylanase; improve; thermophilicity; alkalophilicity;  
 KW thermotolerance; bleach; wood pulp; processing; wheat; maize;  
 KW digestibility-improving animal feed additive; starch production; mutant.

OS Streptomyces lividans.

PN EP828002-A2.

PD 11-MAR-1998.

PF 05-SEP-1997; 97EP-0115412.

PR 09-SEP-1996; 96US-0709912.

PA (CANA) NAT RES COUNCIL CANADA.

PA (NARE-) NAT RES COUNCIL.

PI Ishikawa K, Sung WL, Yaguchi M;

DR WPI: 1998-161100/15.

PT Modified xylanase enzymes - useful for improving wood pulp  
 PT bleaching, etc.

PS Disclosure; Pages 42-43; 84pp; English.

CC AAM60728-44 represent family 11 xylanases. The specification describes a  
 CC method for modifying a family 11 xylanase to improve its  
 CC thermophilicity, alkalophilicity and/or thermotolerance. This method  
 CC comprises modification of amino acids 10, 27 or 29 of Trichoderma reesei  
 CC xylanase II or corresponding aligned amino acids of another family 11  
 CC xylanase, replacement of one or more amino acid sequences in the  
 CC N-terminal region with corresponding aligned sequences from another  
 CC family 11 xylanase to form a chimeric xylanase and/or upstream extension  
 CC of the N terminus by addition of upto 10 amino acids. The modified  
 CC xylanases are useful for improving the bleaching of wood pulp by  
 CC treatment at 55-75 degrees Celsius and pH 7.5-9.0 for 5-180 minutes. They  
 CC might also be useful as digestibility-improving animal feed additives.  
 CC They might also be useful in the processing of wheat or maize for starch  
 CC production.

SQ Sequence 191 AA;

Query Match 51.9%; Score 41; DB 19; Length 191;  
 Best Local Similarity 42.9%; Pred. No. 11;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYQNIQY 14  
 DB 107 SDGCTYDVYQITRY 120

## RESULT 6

ID AAM60281 standard; Protein; 191 AA.

AC AAM60281;

DT 02-SEP-1998 (first entry)

```

XX Modified xylanase C of Streptomyces lividans.
XX
XX Family 11 xylanase; improve; thermophilicity; alkalophilicity;
XX thermotolerance; bleach; wood pulp; processing; wheat; maize;
XX digestibility-improving animal feed additive; starch production; mutant.
XX
XX Synthetic.
XX Streptomyces lividans.
XX
XX Key Location/Qualifiers
XX Misc-difference 15 /label= Y15X
XX /note= "this residue can be Tyr or Phe"
XX
XX EP828002-A2.
XX 11-MAR-1998.
XX
XX 05-SEP-1997; 97EP-0115412.
XX
XX 09-SEP-1996; 96US-0709912.
XX
XX (CANA ) NAT RES COUNCIL CANADA.
XX (NARE-) NAT RES COUNCIL.
XX
XX Ishikawa K, Sung WL, Yaguchi M;
XX WPI; 1998-161100/15.
XX
XX Modified xylanase enzymes - useful for improving wood pulp
XX bleaching, etc.
XX
XX Claim 5; Page -: 84pp; English.
XX
XX The present sequence represents a modified xylanase B of Streptomyces
XX lividans. The specification describes a method for modifying
XX a Family 11 xylanase to improve its thermophilicity, alkalophilicity
XX and/or thermotolerance. This method comprises modification of amino acids
XX 10, 14, 27 or 29 of Trichoderma reesei xylanase II or the corresponding
XX aligned amino acids of another Family 11 xylanase, replacement of one or
XX more amino acid sequences in the N-terminal region with corresponding
XX aligned sequences from another Family 11 xylanase to form a chimeric
XX xylanase and/or upstream extension of the N terminus by addition of up to
XX 10 amino acids. The modified xylanases are useful for improving the
XX bleachability of wood pulp by treatment at 55-75 degrees celsius and
XX pH 7.5-9.0 for 5-180 minutes. They might also be useful as
XX digestibility-improving animal feed additives. They might also be useful
XX in the processing of wheat or maize for starch production.
XX note: this sequence does not appear in the specification; it was created
XX using information provided.
XX
XX Sequence 191 AA;
XX
XX Query Match 51.9%; Score 41; DB 19; Length 191;
XX Best Local Similarity 42.9%; Pred. No. 11;
XX Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 AEGSTXDYQNIQY 14
XX :| | | | | | |
XX 107 SDGGTYDIYQTRX 120
XX
XX RESULT 7
XX AAY9675
XX ID AAY9675 standard; protein; 191 AA.
XX
XX AAY9675;
XX
XX 28-SEP-2000 (first entry)
XX
XX lividans xylanase, Xyn C.
XX
XX

```

```

XX Xylanase; animal feed; digestion efficiency; thermostable;
XX feed pelleting; enzyme; Xyn A; Xyn B; Xyn; Xyn C; Xyn I; Xyn II.
XX
XX Streptomyces lividans.
XX
XX WO200029587-A1.
XX
XX 25-MAY-2000.
XX
XX 16-NOV-1999; 99WO-CA01093.
XX
XX 16-NOV-1998; 98US-0108504.
XX
XX (IOGE-) IOGEN CORP.
XX
XX Sung WL, Tolian JS;
XX WPI; 2000-387799/33.
XX
XX Thermostable xylanases useful for preparing animal feeds especially
XX poultry or swine feed, exhibits optimal activity under physiological
XX conditions.
XX
XX Disclosure; Fig 1; 86pp; English.
XX
XX Xylanase enzymes are added to animal feeds to increase the efficiency of
XX digestion and assimilation of nutrients. Xylanases are preferentially
XX added during the feed pelleting process. To survive the pelleting
XX process and to have optimum activity in the animal, the xylanase needs to
XX have high thermostability, with optimum activity at physiological pH and
XX temperature. The present sequence, xylanase Xyn C, from Streptomyces
XX lividans, is a xylanase Family 11 member. The xylanases of Family
XX 11 have several properties suitable for feed applications, however, they
XX lack the thermostability required to survive food pelleting. The present
XX sequence was used to identify non-conserved residues in Family 11
XX xylanases which could be mutated to introduce desirable properties e.g.
XX thermostability. As a result various thermostable xylanases were
XX identified (AAY9683, AAY9684, AAY9685, AAY9686, AAY9735 and
XX AAY9736) which would be useful for animal feeds, especially poultry and
XX swine feed.
XX
XX Sequence 191 AA;
XX
XX Query Match 51.9%; Score 41; DB 21; Length 191;
XX Best Local Similarity 42.9%; Pred. No. 11;
XX Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 AEGSTXDYQNIQY 14
XX :| | | | | | |
XX 107 SDGGTYDIYQTRX 120
XX
XX RESULT 8
XX AAE18447
XX ID AAE18447 standard; Protein; 191 AA.
XX
XX AAE18447;
XX
XX 16-MAY-2002 (first entry)
XX
XX Streptomyces lividans xylanase, Xyn C.
XX
XX Modified xylanase; thermostability; alkalophilicity; industrial process;
XX pulp manufacture; poultry; swine feed; enzyme; Xyn C.
XX
XX Streptomyces lividans.
XX
XX WO200192487-A2.
XX
XX 06-DEC-2001.
XX
XX 31-MAY-2001; 2001WO-CA00769.
XX
XX

```

	CC	mutated using N-methyl-N'-nitro-N-nitrosocandine and a double mutant beta-1,4-D-glucan glucano hydrolase (endoglucanase)-negative
	CC	and xylanase-negative was selected. Prototyping and
	CC	transformation of the double mutant was carried out using the DNA fragments and the multicopy vector pJ702 to obtain the xylanase C-harboring multicopy clone xlnC. The xylanase enzyme produced by this clone has a higher activity than known xylanasases and can be used for hydrolysing beta-1,4-D-xylosidic linkages in pulp. The treatment of lignocellulose material with the xylanase results in delignification, brightening and viscosity improvement. Further, such treatment may provide more relaxed fibres resulting in an improved performance or a subsequent treatment, such as swelling, CC heating, drainage or chemical bleaching of the pulp, with an overall reduction in energy and chemicals used.
	XX	
SQ	Sequence	200 AA:
	Query Match	51.9%;
	Best Local Similarity	42.9%; Pred. No. 12;
	Matches	6; Conservative 4; Mismatches 4; Indels 0; Gaps 0.
Oy	1 AEGSTDPVONIOY 14	:::   ::   ::
Dd	116 SDGSTDIYQTTRX 129	
	RESULT 10	
ID	AAB48535 standard: Protein; 240 AA.	
AC	AAB48535:	
XX		
DT	05-MAR-2001 (first entry)	
XX		
DE	Streptomyces lividans xylanase C.	
XX		
KW	Bacterial: Bacillus circulans; xylanase; xylanase activity; XA; bleaching agent.	
OS	Streptomyces lividans.	
XX		
PN	WO200068396-A2.	
XX		
PD	16-NOV-2000.	
PX	12-MAY-2000; 2000MO-US13172.	
XX		
PR	12-MAY-1999; 99US-0133714.	
XX		
PA	(XENC-) XENCOR INC.	
XX		
PI	Bentzen JM;	
XX		
DR	WPI, 2000-679800/66.	
XX		
PT	Non naturally occurring XA protein with enhanced thermophilicity, alalophilicity or thermostability relative to the naturally occurring Bacillus circulans xylanase is used in an agent for bleaching pulp -	
XX		
PS	Disclosure; Fig 16D, 114pp; English.	
XX		
CC	The present sequence is given in a specification relating to non naturally occurring xylanase activity (XA) proteins. The XA proteins comprise an amino acid sequence less than 97% identical to a naturally occurring Bacillus circulans xylanase. They are modified to exhibit enhanced thermophilicity, alkalophilicity or thermostability relative to the naturally occurring B. circulans xylanase. They may be used as the active compound in a bleaching agent which is used for bleaching pulp.	
XX		
SQ	Sequence	240 AA;
	Query Match	51.9%; Score 41; DB 21; Length 240;

Best Local Similarity 42.9%; Pred. No. 15;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQY 14  
Db 156 SDGSTRDYQTRK 169

# RESULT 11

ID ABG13264 standard; Protein: 359 AA.  
XX ABG13264:  
AC ABG13264:  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #13255.  
XX  
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB: AAS77451.  
XX  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 43623; 103pp; English.

The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. ABG00010-ABG30377 represent novel human  
diagnostic amino acid sequences. ABG00010-ABG30377 represent novel human  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 359 AA; 51.9%; Score 41; DB 22; Length 359;  
Best Local Similarity 46.7%; Pred. No. 25;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 EGSFXDYQNIQYAG 16  
Db 46 DGFTRFYQRIKYL 60

# RESULT 12

ID ABG08970 standard; Protein: 541 AA.  
XX ABG08970:  
AC ABG08970:  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #8961.  
XX  
XX  
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB: AAS73157.  
XX  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 39329; 103pp; English.

The invention relates to isolated polynucleotide (I) and  
polypeptide (II) sequences. (I) is useful as hybridisation probes,  
polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
and gene mapping, and in recombinant production of (II). The  
polynucleotides are also used in diagnostics as expressed sequence tags  
for identifying expressed genes. (I) is useful in gene therapy techniques  
to restore normal activity of (II) or to treat disease states involving  
(II). (II) is useful for generating antibodies against it, detecting or  
quantitating a polypeptide in tissue, as molecular weight markers and as  
a food supplement. (II) and its binding partners are useful in medical  
imaging of sites expressing (II). (I) and (II) are useful for treating  
disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. ABG00010-ABG30377 represent novel human  
diagnostic amino acid sequences. ABG00010-ABG30377 represent novel human  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 541 AA; 51.9%; Score 41; DB 22; Length 541;  
Best Local Similarity 46.7%; Pred. No. 42;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

DB 230 DGFTEFYORIKYLG 244

## RESULT 13

ABG05639 ID ABG05639 standard; Protein; 1560 AA.

AC ABG05639;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #5630.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dimanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR N-PSDB; AAS69826.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 20; SEQ ID NO 35998; 103pp: English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1560 AA;

Query Match 51.9%; Score 41; DB 22; Length 1560;

Best Local Similarity 46.7%; Pred. No. 1.6e+02;

Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 EGGTXDVYQNIQYAG 16

DB 46 DGFTEFYORIKYLG 60

## RESULT 14

ABB68426 ID ABB68426 standard; Protein; 2659 AA.

AC ABB68426;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 32070.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614450.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li FWD, Myers EM;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL12529.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions

PS Disclosure; SEQ ID NO 32070; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB116175) expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2659 AA;

Query Match 51.9%; Score 41; DB 22; Length 2659;

Best Local Similarity 53.3%; Pred. No. 3.1e+02;

Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 EGGTXDVYQNIQYAG 16

DB 464 EGFQDVYQNLNYCG 498

## RESULT 15

AA71906 ID AA71906 standard; Protein; 313 AA.

AC AA71906;

DT 26-MAR-2001 (first entry)

DE TGS mouse TH2AFL isotype #3.

XX Mouse; TH2AFL isotype; transgenic mouse; TGS; interleukin 9; IL-9;



KW inflammatory response; type 2 helper T-cell; TH; eczema; therapy; atopy;  
 KW allergy; asthma; rhinitis; urticaria; bronchial hyperresponsiveness;  
 KW inflammatory bowel disease; IBD; antiasthmatic; antiinflammatory;  
 KW immunosuppressant.  
 XX  
 OS Mus musculus.  
 XX  
 PN MO20006708-A2.  
 XX  
 PD 09-NOV-2000.  
 XX  
 PF 01-MAY-2000; 2000WO-US11712.  
 XX  
 PR 01-MAY-1999; 99US-0132138.  
 XX  
 PA (MAGA-) MAGAININ PHARM INC.  
 XX  
 PI Louahed J, Dong Q, Levitt RC, Maloy WL, Zhou Y, Nicolaidis NC;  
 DR WPI; 2000-656494/63.  
 DR N-PSDB; AAD01958.  
 XX  
 PT New TH2AF1 genes expressed in association with an inflammatory response  
 PT in airways mediated by type 2 helper T cells are used to treat asthma,  
 PT an asthma-related disorder and inflammatory bowel disease -  
 XX  
 PS Claim 13; Page 81-82; 94pp; English.  
 XX  
 CC The present sequence is a murine TH2AF1 isotype #3  
 CC obtained from small intestine of T65 mouse. Lungs of transgenic  
 CC mouse (T65) which over-expresses IL-9 (Interleukin-9) is used to  
 CC isolate IL-9 induced genes. TH2AF1 gene is expressed in association  
 CC with an inflammatory response in the airways mediated by type 2 helper  
 CC T-cells (TH). The TH2AF1 genes are selectively upregulated by IL-9 and  
 CC are part of the IL-9 signalling pathway. Down regulation of TH2AF1 is  
 CC used in the diagnosis, prevention or treatment of atopic allergy  
 CC including asthma, bronchial hyperresponsiveness, rhinitis, urticaria,  
 CC allergic inflammatory bowel disease (IBD) and various forms of eczema.  
 XX  
 SQ Sequence 313 AA;  
 Query Match 50.6%; Score 40; DB 21; Length 313;  
 Best Local Similarity 57.1%; Pred. No. 33;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYXONIOY 14  
 |||:| | | | |  
 Db 126 AEGATSDDXKNGY 139

Search completed: May 13, 2003, 15:55:37  
 Job time : 77 secs

Tue May 13 16:15:51 2003

us-09-737-297-3.rapb

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 15:57:50 ; Search time 17 Seconds  
(without alignments)  
86.612 Million cell updates/sec

Title: US-09-737-297-3

Sequence: 1 AEGSTXDYQNIQYAG 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PTI\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	77	97.5	16 10 US-09-737-297-3	Sequence 3, Appl1
2	40	50.6	456 10 US-09-815-242-10870	Sequence 10870, A
3	40	50.6	716 10 US-09-845-157-2	Sequence 2, Appl1
4	38.5	48.7	813 10 US-09-764-898-197	Sequence 197, App
5	38	48.1	105 10 US-09-864-761-34387	Sequence 34387, A
6	38	48.1	290 10 US-09-939-980-462	Sequence 462, App
7	38	48.1	313 9 US-10-213-990-72	Sequence 72, Appl1
8	37.5	47.5	146 10 US-09-730-617-74	Sequence 74, Appl1
9	37.5	47.5	149 10 US-09-730-617-68	Sequence 68, Appl1
10	37.5	47.5	149 10 US-09-730-617-69	Sequence 69, Appl1
11	37.5	47.5	149 10 US-09-730-617-70	Sequence 70, Appl1
12	37.5	47.5	157 9 US-10-139-833-17	Sequence 17, Appl1
13	37.5	47.5	157 10 US-09-976-472-2	Sequence 2, Appl1
14	37.5	47.5	170 10 US-09-730-617-10	Sequence 10, Appl1
15	37.5	47.5	170 10 US-09-730-617-77	Sequence 77, Appl1
16	37	46.8	301 9 US-09-738-626-6650	Sequence 6650, Ap
17	36	45.6	455 10 US-09-968-6828-3	Sequence 3, Appl1
18	36	45.6	1456 9 US-09-870-759-95	Sequence 95, App
19	35	44.3	97 9 US-09-925-299-1071	Sequence 1071, Ap

ALIGNMENTS

20	35	44.3	97	9	US-10-102-806-655	Sequence 655, App
21	35	44.3	347	10	US-09-925-299-1071	Sequence 1071, Ap
22	35	44.3	97	10	US-09-770-621-2	Sequence 2, Appl1
23	35	44.3	578	9	US-10-083-357-1344	Sequence 1344, Ap
24	35	44.3	634	10	US-09-883-825-51	Sequence 51, Appl1
25	35	44.3	934	10	US-09-912-697-10	Sequence 10, Appl1
26	35	44.3	934	10	US-09-788-657-19	Sequence 19, Appl1
27	35.5	43.7	1706	2	US-10-024-450-2	Sequence 2, Appl1
28	34	43.0	40	9	US-10-001-835-120	Sequence 120, App
29	34	43.0	64	9	US-10-001-835-176	Sequence 176, App
30	34	43.0	133	9	US-09-790-070A-10	Sequence 10, Appl1
31	34	43.0	146	10	US-09-764-877-2010	Sequence 2010, Ap
32	34	43.0	202	10	US-09-815-242-11841	Sequence 11841, A
33	34	43.0	217	9	US-09-750-070A-11	Sequence 11, Appl1
34	34	43.0	259	10	US-09-815-242-11976	Sequence 11976, A
35	34	43.0	313	9	US-09-992-598-414	Sequence 414, App
36	34	43.0	313	9	US-09-989-293A-414	Sequence 414, App
37	34	43.0	313	9	US-10-063-547-88	Sequence 88, Appl1
38	34	43.0	313	9	US-09-989-735-414	Sequence 414, App
39	34	43.0	313	9	US-09-980-444-414	Sequence 414, App
40	34	43.0	313	9	US-09-989-730-414	Sequence 414, App
41	34	43.0	313	9	US-09-990-436-414	Sequence 414, App
42	34	43.0	313	9	US-09-991-181-414	Sequence 414, App
43	34	43.0	313	9	US-09-993-687-414	Sequence 414, App
44	34	43.0	313	9	US-09-989-734-414	Sequence 414, App
45	34	43.0	313	9	US-09-997-653-414	Sequence 414, App

RESULT 1  
US-09-737-297-3  
Sequence 3, Application US/09737297  
Patent No. US20020072108A1  
GENERAL INFORMATION:  
APPLICANT: Berty, Mark  
APPLICANT: Hill, Philip  
APPLICANT: Laybourne Parry, Johanna  
TITLE OF INVENTION: Processes and Organisms for the Production of Antifreeze F  
FILE REFERENCE: F3247  
CURRENT APPLICATION NUMBER: US/09/737, 297  
CURRENT FILING DATE: 2000-12-15  
PRIOR APPLICATION NUMBER: GB 9929696.4  
PRIOR FILING DATE: 1999-12-15  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 3  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Marinomonas protea  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (6)..(6)  
OTHER INFORMATION: residue 6 is G or V  
US-09-737-297-3  
Query Match 97.5%: Score 77: DB 10: Length 16;  
Best Local Similarity 100.0%: Pred. NO. 1.8e-07;  
Matches 16: Conservative 0: Mismatches 0: Indels 0: Gaps 0;  
OY 1 AEGSTXDYQNIQYAG 16  
Db 1 AEGSTXDYQNIQYAG 16  
RESULT 2  
US-09-815-242-10870  
Sequence 10870, Application US/09815242  
Patent No. US20020061559A1  
GENERAL INFORMATION:

Tue May 13 16:15:51 2003

us-09-737-297-3.rapb

Page 2

```
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: fastseq for Windows Version 4.0
SEQ ID NO 10870
LENGTH: 456
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-815-242-10870
```

```
Query Match
Best Local Similarity 46.7%; Score 40; DB 10; Length 456;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
```

```
OY 2 EGSIXDYVQNIQYAG 16
DB 255 EGSIXDYVQNIQYAG 269

RESULT 3
US-09-845-157-2
Sequence 2, Application US/09845157
Patent No. US20020090618A1
GENERAL INFORMATION:
APPLICANT: Smith, J.
TITLE OF INVENTION: Thermostable Reverse Transcriptases and Uses Thereof
FILE REFERENCE: 0942.5040001
CURRENT APPLICATION NUMBER: US/09/845,157
CURRENT FILING DATE: 2001-05-01
PRIOR APPLICATION NUMBER: US 60//207,196
PRIOR FILING DATE: 2000-05-26
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 716
TYPE: PRT
ORGANISM: M-MLV reverse transcriptase gene
US-09-845-157-2
```

```
Query Match
Best Local Similarity 53.3%; Score 40; DB 10; Length 716;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 1 AEGSTXDVYQNIQYAG 15
DB 606 AEGSTXDVYQNIQYAG 620
```

```
RESULT 4
US-09-764-898-197
Sequence 197, Application US/09764898
Patent No. US20020090673A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PJ201
CURRENT APPLICATION NUMBER: US/09/764,898
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 311
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 197
LENGTH: 813
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-898-197
```

```
Query Match
Best Local Similarity 47.1%; Score 38.5; DB 10; Length 813;
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;
```

```
OY 1 AEGSTXDVYQNIQYAG 16
DB 103 AEGSTXDVYQNIQYAG 119
```

```
RESULT 5
US-09-864-761-34387
Sequence 34387, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aescmca-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,667
```

Tue May 13 16:15:51 2003

us-09-737-297-3.rapb

Page 3

```

; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34387
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007869.1
; OTHER INFORMATION: EXPRESSED IN HPLA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EST_HUMAN HIT: BE086814.1, EVALUOE 2.00e-36
; OTHER INFORMATION: SWISSPROT HIT: P21414, EVALUOE 2.00e-27
US-09-864-761-34387

Query Match      48.1%; Score 38; DB 10; Length 105;
Best Local Similarity 46.7%; Pred. No. 13;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      1 AEGSTXDYVYQNTQYA 15
DB      41 SEGKTVMYNTDSQYA 55
      :||| :| :|||
      :||| :| :|||

RESULT 6
US-09-939-980-462
; Sequence 462, Application US/09939980
; Patent No. US20020082234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
;            Burnham, Martin
;            Hodgson, John
;            Knowles, David
;            Lonetto, Michael
;            Nicholas, Richard
;            Pratt, Julie
;            Reichard, Richard
;            Rosenberg, Martin
;            Ward, Judith
; TITLE OF INVENTION: No. US20020082234A1el Prokaryotic Polynucleotides,
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatiible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,980
; FILING DATE: 27-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/936,165
; FILING DATE: <Unknown>
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 462:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 290 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 462:
US-09-939-980-462

Query Match      48.1%; Score 38; DB 10; Length 290;
Best Local Similarity 40.0%; Pred. No. 40;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY      2 EGSTXDYVYQNTQYAG 16
DB      82 DGTIDLYEGIKETG 96
      :||| :| :|||
      :||| :| :|||

RESULT 7
US-10-213-990-72
; Sequence 72, Application US/10213990
; Publication No. US20030062595A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Bussey, Howard
; APPLICANT: Storms, Reg
; APPLICANT: Roemer, Terry
; TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUNIGATUS ENCODING INDUSTRIAL
; TITLE OF INVENTION: ENZYMES AND METHODS OF USE
; FILE REFERENCE: 10182-019-999
; CURRENT APPLICATION NUMBER: US/10/213,990
; CURRENT FILING DATE: 2002-08-05
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Aspergillus
US-10-213-990-72

Query Match      48.1%; Score 38; DB 9; Length 313;
Best Local Similarity 46.2%; Pred. No. 44;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 AEGSTXDYVYQNTQ 13
DB      143 SDGSTDYDIEHQD 155
      :||| :| :|||
      :||| :| :|||

RESULT 8
US-09-730-617-74
; Sequence 74, Application US/09730617
; Patent No. US20020068279A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine E
; APPLICANT: Prayaga, Sudhirdas K
; APPLICANT: Shinkets, Richard A
; APPLICANT: Rastelli, Luca
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Mezes, Peter S
; TITLE OF INVENTION: No. US20020068279A1el Proteins and Nucleic Acids Encoding
; FILE REFERENCE: 15966-609
; CURRENT APPLICATION NUMBER: US/09/730,617
; CURRENT FILING DATE: 2000-12-05
```

```

      PRIOR APPLICATION NUMBER: 60/169,056
      PRIOR FILING DATE: 1999-12-06
      PRIOR APPLICATION NUMBER: 60/169,886
      PRIOR FILING DATE: 1999-12-09
      PRIOR APPLICATION NUMBER: 60/169,866
      PRIOR FILING DATE: 1999-12-09
      PRIOR APPLICATION NUMBER: 60/170,252
      PRIOR FILING DATE: 1999-12-10
      PRIOR APPLICATION NUMBER: 60/175,740
      PRIOR FILING DATE: 2000-01-12
      NUMBER OF SEQ ID NOS: 100
      SOFTWARE: PatentIn Ver. 2.1
      SEQ ID NO 74
      LENGTH: 146
      TYPE: PRT
      ORGANISM: Homo sapiens
US-09-730-617-74

Query Match      47.5%; Score 37.5; DB 10; Length 146;
Best Local Similarity 53.3%; Pred. No. 23;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 EGSTXDYONICQYAG 16
DB 100 EGST-SVFQSVSYPG 116

RESULT 9
US-09-730-617-68
Sequence 68, Application US/09730617
Patent No. US20020068279A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine E
APPLICANT: Prayaga, Sudhirdas K
APPLICANT: Shimkels, Richard A
APPLICANT: Rastelli, Luca
APPLICANT: Zerhusen, Bryan D
APPLICANT: Mezes, Peter S
TITLE OF INVENTION: No. US20020068279A1e1 Proteins and Nucleic Acids Encoding the Sam
FILE REFERENCE: 15966-609
CURRENT APPLICATION NUMBER: US/09/730,617
PRIOR FILING DATE: 2000-12-05
PRIOR APPLICATION NUMBER: 60/169,056
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 60/169,886
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/169,866
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/170,252
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: 60/175,740
PRIOR FILING DATE: 2000-01-12
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 68
LENGTH: 149
TYPE: PRT
ORGANISM: Homo sapiens
US-09-730-617-68

Query Match      47.5%; Score 37.5; DB 10; Length 149;
Best Local Similarity 53.3%; Pred. No. 24;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 EGSTXDYONICQYAG 16
DB 103 EGST-SVFQSVSYPG 116

RESULT 10
US-09-730-617-69
Sequence 69, Application US/09730617
Patent No. US20020068279A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine E
APPLICANT: Prayaga, Sudhirdas K
APPLICANT: Shimkels, Richard A
APPLICANT: Rastelli, Luca
APPLICANT: Zerhusen, Bryan D
APPLICANT: Mezes, Peter S
TITLE OF INVENTION: No. US20020068279A1e1 Proteins and Nucleic Acids Encoding the Sam
FILE REFERENCE: 15966-609
CURRENT APPLICATION NUMBER: US/09/730,617
PRIOR FILING DATE: 2000-12-05
PRIOR APPLICATION NUMBER: 60/169,056
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 60/169,886
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/169,866
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/170,252
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: 60/175,740
PRIOR FILING DATE: 2000-01-12
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 149
TYPE: PRT
ORGANISM: Homo sapiens
US-09-730-617-70

Query Match      47.5%; Score 37.5; DB 10; Length 149;
Best Local Similarity 53.3%; Pred. No. 24;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 EGSTXDYONICQYAG 16
DB 103 EGST-SVFQSVSYPG 116

RESULT 11
US-09-730-617-70
Sequence 70, Application US/09730617
Patent No. US20020068279A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine E
APPLICANT: Prayaga, Sudhirdas K
APPLICANT: Shimkels, Richard A
APPLICANT: Rastelli, Luca
APPLICANT: Zerhusen, Bryan D
APPLICANT: Mezes, Peter S
TITLE OF INVENTION: No. US20020068279A1e1 Proteins and Nucleic Acids Encoding the Sam
FILE REFERENCE: 15966-609
CURRENT APPLICATION NUMBER: US/09/730,617
PRIOR FILING DATE: 2000-12-05
PRIOR APPLICATION NUMBER: 60/169,056
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 60/169,886
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/169,866
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/170,252
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: 60/175,740
PRIOR FILING DATE: 2000-01-12
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 149
TYPE: PRT
ORGANISM: Homo sapiens
US-09-730-617-70

Query Match      47.5%; Score 37.5; DB 10; Length 149;
Best Local Similarity 53.3%; Pred. No. 24;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 EGSTXDYONICQYAG 16
DB 103 EGST-SVFQSVSYPG 116
```

Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
QY 2 EGSTXDYVONIOTYAG 16  
||||| |:::| |  
Db 103 EGST-SVFQSVSYPG 116

RESULT 12  
US-10-139-833-17  
; Sequence 17, Application US/10139833  
; Publication No. US20030004106A1  
; GENERAL INFORMATION:  
; APPLICANT: Saris, Christian M.  
; APPLICANT: Giles, Jennifer  
; APPLICANT: Mu, Sharon X.  
; APPLICANT: Xia, Min  
; APPLICANT: Bass, Michael B.  
; APPLICANT: Craveiro, Roger  
; TITLE OF INVENTION: Interleukin-1 Receptor Antagonist-Related Molecules and  
; TITLE OF INVENTION: Uses Thereof  
; FILE REFERENCE: 00-1213-E  
; CURRENT APPLICATION NUMBER: US/10/139,833  
; CURRENT FILING DATE: 2002-05-06  
; PRIOR APPLICATION NUMBER: 60/170,191  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/188,053  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: 60/194,521  
; PRIOR FILING DATE: 2000-04-04  
; PRIOR APPLICATION NUMBER: 60/195,910  
; PRIOR FILING DATE: 2000-04-10  
; PRIOR APPLICATION NUMBER: 09/724,583  
; PRIOR FILING DATE: 2000-11-28  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 17  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-139-833-17

Query Match 47.5%; Score 37.5; DB 9; Length 157;  
Best Local Similarity 53.3%; Pred. No. 25;  
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
QY 2 EGSTXDYVONIOTYAG 16  
||||| |:::| |  
Db 111 EGST-SVFQSVSYPG 124

RESULT 13  
US-09-976-472-2  
; Sequence 2, Application US/09976472  
; Patent No. US20020147310A1  
; GENERAL INFORMATION:  
; APPLICANT: SIMS, John E.  
; APPLICANT: RENSCHAW, Blair R.  
; TITLE OF INVENTION: IL-1 ETA DNA AND POLYPEPTIDES  
; FILE REFERENCE: 2932-B  
; CURRENT APPLICATION NUMBER: US/09/976,472  
; CURRENT FILING DATE: 2001-10-11  
; PRIOR APPLICATION NUMBER: PCT/US00/14435  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: 60/162,331  
; PRIOR FILING DATE: 1999-10-29  
; PRIOR APPLICATION NUMBER: 60/135,758  
; PRIOR FILING DATE: 1999-05-25  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-09-976-472-2  
Query Match 47.5%; Score 37.5; DB 10; Length 157;  
Best Local Similarity 53.3%; Pred. No. 25;  
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 EGSTXDYVONIOTYAG 16  
||||| |:::| |  
Db 111 EGST-SVFQSVSYPG 124

RESULT 14  
US-09-730-617-10  
; Sequence 10, Application US/09730617  
; Patent No. US20020068279A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Prayaga, Sudhirdas K  
; APPLICANT: Shinkets, Richard A  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Zernusen, Bryan D  
; APPLICANT: Mezes, Peter S  
; TITLE OF INVENTION: No. US20020068279A1 Proteins, and Nucleic Acids Encoding  
; FILE REFERENCE: 15966-609  
; CURRENT APPLICATION NUMBER: US/09/730,617  
; CURRENT FILING DATE: 2000-12-05  
; PRIOR APPLICATION NUMBER: 60/169,056  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 60/169,886  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/169,866  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/170,252  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/175,740  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-730-617-10

Query Match 47.5%; Score 37.5; DB 10; Length 170;  
Best Local Similarity 53.3%; Pred. No. 28;  
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
QY 2 EGSTXDYVONIOTYAG 16  
||||| |:::| |  
Db 124 EGST-SVFQSVSYPG 137

RESULT 15  
US-09-730-617-77  
; Sequence 77, Application US/09730617  
; Patent No. US20020068279A1  
; GENERAL INFORMATION:  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Prayaga, Sudhirdas K  
; APPLICANT: Shinkets, Richard A  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Zernusen, Bryan D  
; APPLICANT: Mezes, Peter S  
; TITLE OF INVENTION: No. US20020068279A1 Proteins and Nucleic Acids Encoding  
; FILE REFERENCE: 15966-609  
; CURRENT APPLICATION NUMBER: US/09/730,617  
; CURRENT FILING DATE: 2000-12-05  
; PRIOR APPLICATION NUMBER: 60/169,056  
; PRIOR FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 60/169,886  
; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/169,866

Tue May 13 16:15:51 2003

us-09-737-297-3.rapb

Page 6

; PRIOR FILING DATE: 1999-12-09  
; PRIOR APPLICATION NUMBER: 60/170,252  
; PRIOR FILING DATE: 1999-12-10  
; PRIOR APPLICATION NUMBER: 60/175,740  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 77  
; LENGTH: 170  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-730-617-77

Query Match 47.5%; Score 37.5; DB 10; Length 170;  
Best Local Similarity 53.3%; Pred. No. 28;  
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;  
OY 2 EGSTXDYVQNIQYAG 16  
Db 124 EGST-SVFQSVSPG 137

Search completed: May 13, 2003, 16:03:34  
Job time : 19 secs

Tue May 13 16:15:51 2003

us-09-737-297-3.rai

Page 1

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Computec Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 15:54:19 ; Search time 29 Seconds  
(without alignments)  
16.233 Million cell updates/sec

Title: US-09-737-297-3  
Perfect score: 79  
Sequence: 1 AEGSTXDVYQNIQYAG 16

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/ptodata1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata1/iaa/6C\_COMB.pep: \*  
6: /cgn2\_6/ptodata1/iaa/backfillseq1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	41	51.9	191 1	US-08-044-621D-35 Sequence 35, Appl
2	41	51.9	191 2	US-08-709-912-11 Sequence 11, Appl
3	41	51.9	191 3	US-08-047-370-11 Sequence 11, Appl
4	41	51.9	216 1	US-08-315-695-20 Sequence 20, Appl
5	39	49.4	201 4	US-09-311-311C-21 Sequence 21, Appl
6	38.5	48.7	796 3	US-09-005-180A-1 Sequence 1, Appl
7	38	48.1	200 1	US-07-744-570B-2 Sequence 2, Appl
8	38	48.1	290 4	US-08-936-165A-462 Sequence 462, App
9	37	46.8	104 1	US-07-978-692-4 Sequence 4, Appl
10	37	46.8	189 1	US-08-044-621D-33 Sequence 33, Appl
11	37	46.8	189 2	US-08-709-912-12 Sequence 12, Appl
12	37	46.8	189 3	US-09-047-370-12 Sequence 12, Appl
13	37	46.8	200 4	US-08-275-526C-24 Sequence 24, Appl
14	37	46.8	200 4	US-09-076-677-24 Sequence 24, Appl
15	37	46.8	200 4	US-09-073-055-24 Sequence 24, Appl
16	37	46.8	211 1	US-08-575-964-1 Sequence 1, Appl
17	37	46.8	211 2	US-08-563-500-1 Sequence 1, Appl
18	37	46.8	227 4	US-08-275-526C-31 Sequence 31, Appl
19	37	46.8	227 4	US-09-076-677-31 Sequence 31, Appl
20	37	46.8	227 4	US-09-073-055-31 Sequence 31, Appl
21	36	45.6	189 1	US-08-709-912-13 Sequence 13, Appl
22	36	45.6	189 2	US-09-047-370-13 Sequence 13, Appl
23	36	45.6	455 2	US-08-870-827-3 Sequence 3, Appl
24	36	45.6	455 3	US-09-317-179-3 Sequence 3, Appl
25	36	45.6	1455 3	US-08-840-062-3 Sequence 5, Appl
26	36	45.6	3135 1	US-08-323-170B-2 Sequence 2, Appl
27	36	45.6	3135 4	US-08-954-441-2 Sequence 2, Appl

28	35	44.3	184 1	US-08-088-633-6 Sequence 6, Appl
29	35	44.3	184 1	US-08-245-756-6 Sequence 6, Appl
30	35	44.3	184 1	US-08-441-750-6 Sequence 6, Appl
31	35	44.3	184 1	US-08-441-750-6 Sequence 6, Appl
32	35	44.3	184 5	PCT-US82-02521-6 Sequence 6, Appl
33	35	44.3	185 1	US-08-044-621D-36 Sequence 36, Appl
34	35	44.3	185 1	US-08-044-621D-37 Sequence 37, Appl
35	35	44.3	185 1	US-08-709-912-3 Sequence 3, Appl
36	35	44.3	185 1	US-08-709-912-3 Sequence 3, Appl
37	35	44.3	185 2	US-09-047-370-5 Sequence 5, Appl
38	35	44.3	185 2	US-09-047-370-5 Sequence 5, Appl
39	35	44.3	208 1	US-08-315-695-21 Sequence 21, Appl
40	35	44.3	213 1	US-08-104-445-3 Sequence 3, Appl
41	35	44.3	273 4	US-08-928-218-10 Sequence 10, Appl
42	35	44.3	344 2	US-08-468-812-2 Sequence 2, Appl
43	35	44.3	344 4	US-08-590-563-2 Sequence 2, Appl
44	35	44.3	353 3	US-08-966-318-3 Sequence 3, Appl
45	35	44.3	353 4	US-09-216-619-3 Sequence 3, Appl

#### ALIGNMENTS

RESULT 1  
US-08-044-621D-35  
Sequence 35 Application US/08044621D  
Patent No. 5405769  
GENERAL INFORMATION:  
APPLICANT: Warren W. Makarchuk  
APPLICANT: Wang L. Sung  
APPLICANT: Makoto Yaguchi  
APPLICANT: Robert L. Campbell  
APPLICANT: David R. Rose  
TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS  
TITLE OF INVENTION: OF A LOW MOLECULAR MASS XYLANASE  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Gowling, Strathly & Henderson  
STREET: Suite 2600, 160 Elgin Street  
CITY: Ottawa  
STATE: Ontario  
COUNTRY: Canada  
ZIP: K1P 1C3  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 in., 360KB storage  
COMPUTER: IBM PC  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/044,621D  
FILING DATE: April 8, 1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Judy A. Erratt  
REGISTRATION NUMBER: 34,076  
REFERENCE/DOCKET NUMBER: 08-863796  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 613-786-0199  
TELEFAX: 613-563-9869  
TELEX:  
INFORMATION FOR SEQ. ID NO.: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 191  
TYPE: Amino Acid  
STRANDEDNESS: No, 5405769 Relevant  
TOPOLOGY: Linear  
MOLECULE TYPE:  
DESCRIPTION: Protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO



Tue May 13 16:15:51 2003

us-09-737-297-3.rai

Page 2

```
FRAGMENT TYPE: NO
ORIGINAL SOURCE:
ORGANISM: Streptomyces lividans
STRAIN: Streptomyces lividans, Xln C
IMMEDIATE SOURCE:
POSITION IN GENOME:
FEATURE:
PUBLICATION INFORMATION:
AUTHORS: Shareck, F., Roy, C., Yaguchi, M.,
AUTHORS: Morosoli, R. & Kluempfel, D.
TITLE:
JOURNAL: Gene
VOLUME: 107
ISSUE:
PAGES: 75-82
DATE: 1991
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-044-621D-35

Query Match 51.9%; Score 41; DB 1; Length 191;
Best Local Similarity 42.9%; Pred. No. 3.1;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYONIQY 14
DB 107 SDGSTDYDYYOTTR 120

RESULT 2
US-08-709-912-11
Sequence 11, Application US/08709912
Patent No. 5759840
GENERAL INFORMATION:
APPLICANT: Sung Dr., Wing L
APPLICANT: Yaguchi Dr., Makoto
APPLICANT: Ishikawa Dr., Kazuhiko
TITLE OF INVENTION: Modification of xylanase to improve
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and
TITLE OF INVENTION: Thermostability
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto
STREET: 277 Park Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10172-0194
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/709,912
FILING DATE: 09-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Olsen Mr., Warren E
REGISTRATION NUMBER: 27290
REFERENCE/DOCKET NUMBER: 1039,2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-2982
TELEFAX: (212) 758-2982
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
```

```
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Streptomyces lividans
STRAIN: Xln C
PUBLICATION INFORMATION:
AUTHORS: Shareck, F
AUTHORS: Roy, C
AUTHORS: Yaguchi, M
AUTHORS: Morosoli, R
AUTHORS: Kluempfel, D
JOURNAL: Gene
VOLUME: 107
PAGES: 75-82
DATE: 1991
US-08-709-912-11

Query Match 51.9%; Score 41; DB 1; Length 191;
Best Local Similarity 42.9%; Pred. No. 3.1;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYONIQY 14
DB 107 SDGSTDYDYYOTTR 120

RESULT 3
US-09-047-370-11
Sequence 11, Application US/09047370
Patent No. 5866408
GENERAL INFORMATION:
APPLICANT: Sung Dr., Wing L
APPLICANT: Yaguchi Dr., Makoto
APPLICANT: Ishikawa Dr., Kazuhiko
TITLE OF INVENTION: Modification of xylanase to improve
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and
TITLE OF INVENTION: Thermostability
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto
STREET: 277 Park Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10172-0194
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/047,370
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/709,912
FILING DATE: 09-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Olsen Mr., Warren E
REGISTRATION NUMBER: 27290
REFERENCE/DOCKET NUMBER: 1039,2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-2982
TELEFAX: (212) 758-2982
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
```

Tue May 13 16:15:51 2003

us-09-737-297-3.ral

Page 3

ANTI-SENSE: NO  
FRAGMENT TYPE: Internal  
ORIGINAL SOURCE:  
ORGANISM: Streptomyces lividans  
STRAIN: Xln C  
PUBLICATION INFORMATION:  
AUTHORS: Shareck, F  
AUTHORS: Roy, C  
AUTHORS: Yaguchi, M  
AUTHORS: Morosoli, R  
AUTHORS: Kluepfel, D  
JOURNAL: Gene  
VOLUME: 107  
PAGES: 75-82  
DATE: 1991  
US-09-047-370-11

Query Match 51.9%; Score 41; DB 2; Length 191;  
Best Local Similarity 42.9%; Pred. No. 3.1;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVYONIQY 14  
DB 107 SDGSTDYVYOTTRY 120

RESULT 4  
US-08-315-695-20  
Sequence 20, Application US/08315695  
Patent No. 5591619

GENERAL INFORMATION:  
APPLICANT: Li, Xin-Liang  
APPLICANT: Jungdahl, Lars G.  
TITLE OF INVENTION: Aureobasidium Pullulans Xylanase, Gene  
TITLE OF INVENTION: and Signal Sequence  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Wanner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: CO  
COUNTRY: US  
ZIP: 80303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/315,695  
FILING DATE: 30-SEP-1994

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Wanner, Ellen P.  
REGISTRATION NUMBER: 28,547  
REFERENCE/DOCKET NUMBER: 55-94  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 216 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-315-695-20

Query Match 51.9%; Score 41; DB 1; Length 216;  
Best Local Similarity 42.9%; Pred. No. 3.6;

Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 1 AEGSTXDYVYONIQY 14  
DB 133 SDGSTDYVYOTTRY 146

RESULT 5  
US-09-311-311C-21

Sequence 21, Application US/09311311C  
Patent No. 6358738  
GENERAL INFORMATION:  
APPLICANT: Erikson, et al.  
TITLE OF INVENTION: POLO BOX THERAPEUTIC COMPOSITIONS,  
METHODS, AND USES THEREFOR  
FILE REFERENCE: 1874/117  
CURRENT APPLICATION NUMBER: US/09/311,311C  
CURRENT FILING DATE: 1999-05-13  
PRIOR APPLICATION NUMBER: US 60/085,296  
PRIOR FILING DATE: 1998-05-13  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 21

LENGTH: 201  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (376)...(576)  
OTHER INFORMATION: Polo protein C-terminal portion  
US-09-311-311C-21

Query Match 49.4%; Score 39; DB 4; Length 201;  
Best Local Similarity 53.8%; Pred. No. 7.9;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GSTXDYVYONIQY 15  
DB 176 GVSXDYVYONIQY 188

RESULT 6  
US-09-005-180A-1  
Sequence 1, Application US/09005180A  
Patent No. 6124446  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi

TITLE OF INVENTION: HUMAN VPS35/MEM3-RELATED PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/005,180A  
FILING DATE: Filed January 8, 1998  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

Tue May 13 16:15:51 2003

us-09-737-297-3.rai

Page 4

REFERENCE/DOCKET NUMBER: PF-0457 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 796 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: LUNGTOT08  
CLONE: 2641812  
US-09-005-180A-1

Query Match 48.1%; Score 38.5; DB 3; Length 796;  
Best Local Similarity 47.1%; Pred. No. 50;  
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 AEG-STXDYVQNIQYAG 16  
Db 86 AKGRKADLYELVQYAG 102

RESULT 7  
US-07-744-570B-2  
Sequence 2, Application US/07744570B  
Patent No. 5202249  
GENERAL INFORMATION:  
APPLICANT: Kluepfel, D.  
APPLICANT: Morosoli, R.  
APPLICANT: Shareck, F.  
TITLE OF INVENTION: Xylanase for Biobleaching  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Michael J. Bradley  
STREET: 1200 South 47th Street  
STREET: Box Number 4023  
CITY: Richmond  
STATE: California  
COUNTRY: United States  
ZIP: 94804-0023  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44MB storage  
COMPUTER: IBM  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07744,570B  
FILING DATE: 19910813  
CLASSIFICATION: 435  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: Single strand  
TOPOLOGY: Circular  
US-07-744-570B-2

Query Match 48.1%; Score 38; DB 1; Length 200;  
Best Local Similarity 35.7%; Pred. No. 12;  
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVQNIQY 14  
Db 116 SDGRTYDIYETTRY 129

RESULT 8  
US-08-936-165A-462  
Sequence 462, Application US/08936165A  
Patent No. 6348582

GENERAL INFORMATION:  
APPLICANT: Black, Michael  
APPLICANT: Burnham, Martin  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Lonetto, Michael  
APPLICANT: Nicholas, Richard  
APPLICANT: Pratt, Julie  
APPLICANT: Reichard, Richard  
APPLICANT: Rosenberg, Martin  
APPLICANT: Ward, Judith  
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,  
TITLE OF INVENTION: Polypeptides and Their Uses  
NUMBER OF SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/936,165A  
FILING DATE: 24-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,032  
FILING DATE: 24-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 462:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 290 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-936-165A-462

Query Match 48.1%; Score 38; DB 4; Length 290;  
Best Local Similarity 40.0%; Pred. No. 19;  
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 EGSITXDYVQNIQYAG 16  
Db 82 DGTIDYEGIKETG 96

RESULT 9  
US-07-978-692-4  
Sequence 4, Application US/07978692  
Patent No. 5457045  
GENERAL INFORMATION:  
APPLICANT: Anker, Lisbeth  
APPLICANT: Bisgaard-Frantzen, Henrik  
APPLICANT: Halkier, Torben  
TITLE OF INVENTION: ENZYMES WITH XYLANOXYTIC ACTIVITY  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 54570450 No. 5457045disk of No. 5457045th America, Inc.  
STREET: 405 Lexington Avenue, Suite 6200  
CITY: New York

Tue May 13 16:15:51 2003

us-09-737-297-3.rai

Page 5

STATE: New York  
COUNTRY: United States of America  
ZIP: 10174-6201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/978,692  
FILING DATE: 19930202  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/00242  
FILING DATE: 23-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 0309/91  
FILING DATE: 22-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK90/00220  
FILING DATE: 24-AUG-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Lambiris, Elias J.  
REGISTRATION NUMBER: 33728  
REFERENCE/DOCKET NUMBER: 3599.204-US  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 104 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
FRAGMENT TYPE: C-terminal  
ORIGINAL SOURCE:  
ORGANISM: Bacillus pumilus  
STRAIN: DSM 6124  
US-07-978-692-4

Query Match 46.8%; Score 37; DB 1; Length 104;  
Best Local Similarity 38.5%; Pred. No. 8.7;  
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQ 13  
DB 20 ADGGTYDIETLR 32

RESULT 10  
US-08-044-621D-33  
Sequence 33 Application US/08044621D  
Patent No. 5405769  
GENERAL INFORMATION:  
APPLICANT: Warren W. Makarchuk  
APPLICANT: Wang L. Sung  
APPLICANT: Makoto Yaguchi  
APPLICANT: Robert L. Campbell  
APPLICANT: David R. Rose  
TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS  
TITLE OF INVENTION: OF A LOW MOLECULAR MASS XYLANASE  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Gowling, Strathy & Henderson  
STREET: Suite 2600, 160 Elgin Street  
CITY: Ottawa  
STATE: Ontario  
COUNTRY: Canada  
ZIP: K1P 1C3  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 in., 360KB storage  
COMPUTER: IBM PC  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/044,621D  
FILING DATE: April 8, 1993

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Judy A. Erialt  
REGISTRATION NUMBER: 34,076  
REFERENCE/DOCKET NUMBER: 08-863796  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 613-786-0199  
TELEFAX: 613-563-9869  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 189  
TYPE: Amino Acid  
STRANDEDNESS: No. 5405769 Relevant  
TOPOLOGY: linear  
MOLECULE TYPE:  
DESCRIPTION: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptomyces sp. #36a  
STRAIN: Streptomyces sp. #36a  
IMMEDIATE SOURCE:  
POSITION IN GENOME:  
FEATURE:  
PUBLICATION INFORMATION:  
AUTHORS: Nagashima M., Okumoto Y. & Okanishi M.  
TITLE: Trends in Actinomycetologia  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES: 91-96  
DATE: 1989  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-044-621D-33

Query Match 46.8%; Score 37; DB 1; Length 189;  
Best Local Similarity 35.7%; Pred. No. 18;  
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQ 14  
DB 107 SDGGTYDIYKTRTY 120

RESULT 11  
US-08-709-912-12  
Sequence 12 Application US/08709912  
Patent No. 5759840  
GENERAL INFORMATION:  
APPLICANT: Sung Dr., Wang L.  
APPLICANT: Yaguchi Dr., Makoto  
APPLICANT: Ishikawa Dr., Kazuhiko  
TITLE OF INVENTION: Modification of xylanase to improve  
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto  
STREET: 277 Park Ave.  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10173-0194  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

Tue May 13 16:15:51 2003

us-09-737-297-3.ra1

Page 6

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/709,912
FILING DATE: 09-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Olsen Mr, Warren E
REGISTRATION NUMBER: 27290
REFERENCE/DOCKET NUMBER: 1039,2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-2400
TELEFAX: (212) 758-2982
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 189 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Streptomyces sp. 36a
PUBLICATION INFORMATION:
AUTHORS: Nagashima, M
AUTHORS: Okumoto, Y
AUTHORS: Okanishi, M
JOURNAL: Trends in Actinomycetologia
PAGES: 91-96
DATE: 1989
US-08-709-912-12

Query Match
Best Local Similarity 35.7%; Score 37; DB 1; Length 189;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYQNIQY 14
DB 107 SDGGTYDIYKTRTY 120

RESULT 12
US-09-047-370-12
Sequence 12, Application US/09047370
Patent No. 5866408
GENERAL INFORMATION:
APPLICANT: Sung Dr., Wing L
APPLICANT: Yaguchi Dr., Makoto
APPLICANT: Iehikawa Dr., Kazuhiko
TITLE OF INVENTION: Modification of xylanase to improve
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and
TITLE OF INVENTION: Thermostability
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSER: Fitzpatrick, Cella, Harper, and Scinto
STREET: 277 Park Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10172-0194
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/047,370
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
```

```

APPLICATION NUMBER: US 08/709,912
FILING DATE: 09-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Olsen Mr, Warren E
REGISTRATION NUMBER: 27290
REFERENCE/DOCKET NUMBER: 1039,2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-2400
TELEFAX: (212) 758-2982
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 189 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Streptomyces sp. 36a
PUBLICATION INFORMATION:
AUTHORS: Nagashima, M
AUTHORS: Okumoto, Y
AUTHORS: Okanishi, M
JOURNAL: Trends in Actinomycetologia
PAGES: 91-96
DATE: 1989
US-09-047-370-12

Query Match
Best Local Similarity 35.7%; Score 37; DB 2; Length 189;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYQNIQY 14
DB 107 SDGGTYDIYKTRTY 120

RESULT 13
US-08-275-526C-24
Sequence 24, Application US/08275526C
Patent No. 6180382
GENERAL INFORMATION:
APPLICANT: DE BOYL, ERIC
APPLICANT: LAHAYE, ANDR E
APPLICANT: LEDOUX, PIERRE
APPLICANT: AMORY, ANTOINE
APPLICANT: DETROZ, REN
APPLICANT: ANDRE, CHRISTOPHE
APPLICANT: VETTER, ROMAN
TITLE OF INVENTION: XYLANASE DERIVED FROM A BACILLUS SPECIES,
TITLE OF INVENTION: EXPRESSION VECTORS FOR SUCH XYLANASE AND
TITLE OF INVENTION: OTHER PROTEINS, HOST ORGANISMS THEREFOR AND
TITLE OF INVENTION: USE THEREOF
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSER: WILLIAM BRINKS HOPFER GILSON & LIONE, P.C.
STREET: 2000 K St., N.W., Suite 200
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/275,526C
FILING DATE: 15-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
```

NAME: Gadiano, Wilhem F.  
REGISTRATION NUMBER: 37,136  
REFERENCE/DOCKET NUMBER: 4121-49  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 429-0625  
TELEFAX: (202) 293-0625  
TELEX: 650 383 5605  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal fragment  
ORIGINAL SOURCE:  
ORGANISM: Bacillus pumilus  
STRAIN: PRL B12  
US-09-275-526C-24

Query Match 46.8%; Score 37; DB 4; Length 200;  
Best Local Similarity 38.5%; Pred. No. 19;  
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 AEGSTXDYXONIQ 13  
DB 115 ADGSTDYIYETLR 127

RESULT 14  
US-09-076-677-24  
Sequence 24, Application US/09076677  
Patent No. 6423523  
GENERAL INFORMATION:  
APPLICANT: DE BUYL, ERIC  
LAHAYE, ANDRÉE  
LEDoux, PIERRE  
AMORY, ANTOINE  
DETROZ, RENE  
ANDRE, CHRISTOPHE  
VETTER, ROMAN  
TITLE OF INVENTION: XYLANASE DERIVED FROM A BACILLUS SPECIES,  
EXPRESSION VECTORS FOR SUCH XYLANASE AND  
OTHER PROTEINS, HOST ORGANISMS THEREFOR AND  
USE THEREOF  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.  
STREET: 2000 K St., N.W., Suite 200  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/076,677  
FILING DATE: 12-MAY-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/275,526  
FILING DATE: 15-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gadiano, Wilhem F.  
REGISTRATION NUMBER: 37,136  
REFERENCE/DOCKET NUMBER: 4121-49  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 429-0625  
TELEFAX: (202) 293-0625  
TELEX: 650 383 5605

INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal fragment  
ORIGINAL SOURCE:  
ORGANISM: Bacillus pumilus  
STRAIN: PRL B12  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-09-076-677-24  
Query Match 46.8%; Score 37; DB 4; Length 200;  
Best Local Similarity 38.5%; Pred. No. 19;  
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 AEGSTXDYXONIQ 13  
DB 115 ADGSTDYIYETLR 127

RESULT 15  
US-09-073-055-24  
Sequence 24, Application US/09073055  
Patent No. 6426211  
GENERAL INFORMATION:  
APPLICANT: DE BUYL, ERIC  
LAHAYE, ANDRÉE  
LEDoux, PIERRE  
AMORY, ANTOINE  
DETROZ, RENE  
ANDRE, CHRISTOPHE  
VETTER, ROMAN  
TITLE OF INVENTION: XYLANASE DERIVED FROM A BACILLUS SPECIES,  
EXPRESSION VECTORS FOR SUCH XYLANASE AND  
OTHER PROTEINS, HOST ORGANISMS THEREFOR AND  
USE THEREOF  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WILLIAM BRINKS HOFER GILSON & LIONE, P.C.  
STREET: 2000 K St., N.W., Suite 200  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/073,055  
FILING DATE: 05-MAY-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/275,526  
FILING DATE: 15-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gadiano, Wilhem F.  
REGISTRATION NUMBER: 37,136  
REFERENCE/DOCKET NUMBER: 4121-49  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 429-0625  
TELEFAX: (202) 293-0625  
TELEX: 650 383 5605  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 200 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

```
MOLECULE TYPE: protein
FRAGMENT TYPE: internal fragment
ORIGINAL SOURCE:
ORGANISM: Bacillus pumilus
STRAIN: PRL B12
SEQUENCE DESCRIPTION: SEQ ID NO: 24
US-09-073-055-24

Query Match      46.8%; Score 37; DB 4; Length 200;
Best Local Similarity 38.5%; Pred. No. 19;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 AEGSTXDYONIQ 13
   ||| |||:|:|
Db 115 ADGQYDIYEFLR 127

Search completed: May 13, 2003, 15:59:07
Job time : 30 secs
```

---

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 15:53:19 ; Search time 44 Seconds  
(without alignments)  
34.958 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79  
Sequence: 1 AEGSTXDYQNIQYAG 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.5	56.3	455	1 VHIHMJ	nucleocapsid prote
2	43	54.4	380	2 A86712	hypothetical prote
3	42	53.2	380	2 H95129	glucose-1-phosphat
4	42	53.2	380	2 P98000	glucose-1-phosphat
5	42	53.2	512	2 S50599	SMC1 protein - yea
6	41	51.9	169	1 B64772	yejO protein - Esc
7	41	51.9	169	2 H90688	hypothetical prote
8	41	51.9	169	2 D85539	hypothetical prote
9	41	51.9	240	1 U50591	endo-1,4-beta-xyla
10	41	51.9	241	2 T37005	endo-1,4-beta-xyla
11	40	50.6	455	2 D81809	hypothetical prote
12	40	50.6	724	2 T33332	hypothetical prote
13	39.5	50.0	60	2 D98032	type II site-speci
14	39	49.4	313	2 A55122	metalloproteinase
15	39	49.4	576	2 S22127	protein kinase pol
16	39	49.4	623	2 AH1209	B. subtilis minor
17	39	49.4	771	2 S45048	capsid protein - h
18	39	49.4	829	2 T19494	hypothetical prote
19	38.5	48.7	545	1 J50533	amidase (EC 3.5.1.
20	38.5	48.7	796	2 UC7516	vesicle protein so
21	38	48.1	149	2 S40261	hypothetical prote
22	38	48.1	229	2 D90958	hypothetical prote
23	38	48.1	229	2 E85806	hypothetical prote
24	38	48.1	250	2 C56339	glucose-1-phosphat
25	38	48.1	284	2 G71192	probable cobalt tr
26	38	48.1	323	2 G81441	nifu protein homol
27	38	48.1	497	2 E90061	hypothetical prote
28	38	48.1	526	2 AG0200	probable methyl-ac
29	38	48.1	643	2 T03518	hypothetical prote

30	38	48.1	785	2 T11719	probable vacuolar
31	38	48.1	921	2 G83902	hypothetical prote
32	38	48.1	1051	2 A38373	ubiquitin-protein
33	38	48.1	1146	2 S40311	integrin - fruit f
34	38	48.1	1374	2 D72593	hypothetical prote
35	37.5	47.5	455	1 D45340	nucleocapsid prote
36	37	46.8	159	2 A91445	single-stranded DN
37	37	46.8	175	2 S75258	hypothetical prote
38	37	46.8	210	2 C83762	endo-1,4-beta-xyla
39	37	46.8	240	1 S47512	endo-1,4-beta-xyla
40	37	46.8	312	2 D89763	conserved hypotet
41	37	46.8	315	2 A96782	unknown protein F2
42	37	46.8	327	2 T44111	conserved hypotet
43	37	46.8	372	2 AD2111	twitching motility
44	37	46.8	423	1 H40382	isocitrate dehydro
45	37	46.8	447	2 H72089	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

VHIHMJ nucleocapsid protein - murine hepatitis virus (strain JHM)

C:Species: murine hepatitis virus, MHV

C:Date: 28-Aug-1985 #sequence\_revision 28-Aug-1985 #text\_change 24-Sep-1999

C:Accession: A04024

R:Skinner, M.A.; Siddell, S.G.

Nucleic Acids Res. 11, 5045-5054, 1983

A:Title: Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocaps

A:Reference number: A04024; MUID:83272950; PMID:6308569

A:Accession: A04024

A:Molecule type: genomic RNA

A:Residues: 1-455 <SKI>

A:Cross-references: GB:X00990; GB:K00757; GB:M25875; NID:958972; PIDN:CAA25497.1;

C:Comment: This genome is negative, linear, single-stranded RNA.

C:Superfamily: coronavirus nucleocapsid protein

C:Keywords: glycoprotein; nucleocapsid

F:410/423/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 56.3%; Score 44.5; DB 1; Length 455;  
Best Local Similarity 56.2%; Pred. No. 3.5;  
Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 AEGSTXDYQNIQYAG 16  
DB 340 ADGPTKDYAE-LQYSG 354

##### RESULT 2

A86712 hypothetical protein glgc [imported] - Lactococcus lactis subsp. lactis (strain I

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001

C:Accession: A86712

R:Boletín, A.; Wincker, P.; Manguer, S.; Jallion, O.; Malarme, K.; Weissenbach, J.

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus la

A:Reference number: A86712; MUID:21235186; PMID:11337471

A:Accession: A86712

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-380 <STO>

A:Cross-references: GB:AE005176; PID:912723607; PIDN:AAK04795.1; GSPDB:GNO0146

A:Experimental source: strain IL1403

A:Gene: glgc

C:Genetics:

C:Superfamily: glucose-1-phosphate adenylyltransferase

Query Match 54.4%; Score 43; DB 2; Length 380;  
Best Local Similarity 61.5%; Pred. No. 5.3;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;



QY 2 EGSTXDYVYONIQY 14  
 ||: |||||  
 DB 100 EGSHAYVONIDY 112

RESULT 3  
 H95129  
 C:Species: Streptococcus pneumoniae  
 C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 24-Aug-2001  
 C:Accession: H95129  
 R:Retellin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held  
 on, J.D.; Umeyama, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
 nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001  
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
 A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
 A:Reference number: A95000; PMID:21357209; PMID:11463916  
 A:Accession: H95129  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-380 <NUP>  
 A:Cross-references: GB:AE005672; PIDN:AAK75233.1; PID:g14972599; GSPDB:GN00164; TIGR:SP4  
 A:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SP1122  
 C:Superfamily: glucose-1-phosphate adenylyltransferase

Query Match 53.2%; Score 42; DB 2; Length 380;  
 Best Local Similarity 53.8%; Pred. No. 8.1;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVYONIQY 14  
 ||: |||||  
 DB 100 EGSHAYVONIDY 112

RESULT 4  
 F98000  
 C:Species: Streptococcus pneumoniae  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
 C:Accession: F98000  
 R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; H  
 e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
 y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
 A:Reference number: A97872; PMID:21429245; PMID:11544234  
 A:Accession: F98000  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-380 <NUP>  
 A:Cross-references: GB:AE007317; PIDN:AAK99834.1; PID:g15458648; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: glpC  
 C:Superfamily: glucose-1-phosphate adenylyltransferase  
 C:Keywords: nucleotidyltransferase

Query Match 53.2%; Score 42; DB 2; Length 380;  
 Best Local Similarity 53.8%; Pred. No. 8.1;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVYONIQY 14  
 ||: |||||  
 DB 100 EGSHAYVONIDY 112

RESULT 5  
 S50599  
 SHCI protein - yeast (*Saccharomyces cerevisiae*)  
 N:Alternate names: protein YER096w

C:Species: *Saccharomyces cerevisiae*  
 C:Date: 28-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 19-Apr-2002  
 C:Accession: S50599; B38215  
 R:Dieckrich, F.S.  
 submitted to the EMBL Data Library, December 1994  
 A:Description: The sequence of *S. cerevisiae* cosmid 9747, 8198, 9781, and lambda c  
 A:Reference number: S50436  
 A:Accession: S50599  
 A:Molecule type: DNA  
 A:Residues: 1-512 <DIE>  
 A:Cross-references: EMBL:U18839; NID:9603313; PID:9603334; MIPS:YER096w  
 R:Shiohara, A.; Ogawa, H.; Ogawa, T.  
 Cell 69, 457-470, 1992  
 A:Title: Rad51 protein involved in repair and recombination in *S. cerevisiae* is a  
 A:Reference number: A38215; PMID:92257587; PMID:11581961  
 A:Accession: B38215  
 A:Molecule type: DNA  
 A:Residues: 228, 230, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245,  
 A:Cross-references: EMBL:D10023; NID:9218468; PID:41001378; PID:g2160326  
 A:Note: sequence extracted from NCBI backbone (NCBIN:101730, NCBI:P.101737)  
 C:Genetics:  
 A:Gene: SGD:SHCI; SHCI  
 A:Cross-references: SGD:S0000898  
 A:Map position: 5R

Query Match 53.2%; Score 42; DB 2; Length 512;  
 Best Local Similarity 53.6%; Pred. No. 12;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TXDVYONIQYA 15  
 ||: |||||  
 DB 255 TQDIYSNFGYA 265

RESULT 6  
 B64772  
 YajQ protein - *Escherichia coli* (strain K-12)  
 N:Alternate names: protein b0426  
 C:Species: *Escherichia coli*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Mar-2002  
 C:Accession: B64772  
 R:Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley,  
 A.; Rose, D.J.; Mau, B.; Zhao, Y.  
 Science 277, 1433-1462, 1997  
 A:Title: The complete genome sequence of *Escherichia coli* K-12.  
 A:Reference number: A64720; PMID:97426617; PMID:9278503  
 A:Accession: B64772  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-169 <BLAT>  
 A:Cross-references: GB:AE000149; GB:U00096; NID:g1786628; PIDN:AACT3529.1; PID:g17  
 A:Experimental source: strain K-12, substrain MG1655  
 C:Genetics:  
 A:Gene: YajQ  
 C:Superfamily: hypothetical protein b0426

Query Match 51.9%; Score 41; DB 1; Length 169;  
 Best Local Similarity 53.3%; Pred. No. 4.9;  
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVYONIQYAG 16  
 ||: |||||  
 DB 81 EGSSLDYVENVIVHSG 95

RESULT 7  
 H90688  
 hypothetical protein EC0480 (imported) - *Escherichia coli* (strain O157:H7, substr  
 C:Species: *Escherichia coli*  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C:Accession: H90688  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han  
 gasawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

```

A:Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans
A:Pathway: xylan degradation
C:Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology
C:Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradat
F:1-49/Domain: signal sequence #status predicted <SIG>
F:50-240/Product: endo-1,4-beta-xylanase C #status experimental <MAT>
F:62-229/Domain: endo-1,4-beta-xylanase homology <XYL>
F:134,226/Active site: Glu #status predicted

Query Match          51.9%:  Score 41;  DB 1;  Length 240;
Best Local Similarity 42.9%:  Pred. No. 7.3;
Matches 6;  Conservative 4;  Mismatches 4;  Indels 0;  Gaps 0;

QY      1 AEGSTXDYQNIQY 14
        ||| ||| |||
Db       156 SDGCTYDIYQTRX 169

RESULT 10
T37005
endo-1,4-beta-xylanase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C:Accession: T37005
C:Rollver, K.; Harris, D.; Bentley, S.D.; Parthill, J.; Barrell, B.G.; Rajandream,
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21618
A:Accession: T37005
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1,241 <OLI>
A:Cross-references: EMBL:AL109949; PIDD:GAB52919.1; GSFDB:GN00070; SCOEDB:SCJ1.3
C:Genetics:
A:Experimental source: strain A3(2)
C:Superfamily: endo-1,4-beta-xylanase; endo-1,4-beta-xylanase homology

Query Match          51.9%:  Score 41;  DB 2;  Length 241;
Best Local Similarity 42.9%:  Pred. No. 7.3;
Matches 6;  Conservative 4;  Mismatches 4;  Indels 0;  Gaps 0;

QY      1 AEGSTXDYQNIQY 14
        ||| ||| |||
Db       157 SDGCTYDIYQTRX 170

RESULT 11
DB1809
hypothetical protein NMA1831 [imported] - Neisseria meningitidis (strain Z2491 se
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: DB1809
R:Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.;
Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajc
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: DB1809
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-455 <PAR>
A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380377; PIDD:GAB5056.1; PID:
A:Experimental source: strain Z2491
C:Genetics:
A:Gene: NMA1831

Query Match          50.6%:  Score 40;  DB 2;  Length 455;
Best Local Similarity 61.5%:  Pred. No. 24;
Matches 8;  Conservative 1;  Mismatches 4;  Indels 0;  Gaps 0;

QY      4 STDYQNIQYAG 16
        |||| | ||
Db       352 SADDYQRIEYAG 364

```

## RESULT 12

T23322

hypothetical protein K04G2.3 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*  
 C:Date: 19-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
 C:Accession: T23322

R:Gardner, A.

submitted to the EMBL Data Library, July 1996

A:Reference number: Z19727

A:Accession: T23322

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-724 &lt;MIL&gt;

A:Cross-references: EMBL:275712; PIDN:CA00040.1; GSPDB:GN00019; CESP:K04G2.3

A:Experimental source: clone K04G2

C:Genetics:

A:Gene: CESP:K04G2.3

A:Map position: 1 347/3; 617/1, 697/1

A:Insertions: 72/3; 347/3; 617/1, 697/1

C:Superfamily: transitional endoplasmic reticulum ATPase; FtsH/Sec18/CDC48-type ATP-bind

Query Match

Best Local Similarity 50.6%; Score 40; DB 2; Length 724;

Best Local Similarity 66.7%; Pred. No. 40;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 AEGSTDYVONI 12

DB 230 AEGSTDYVONI 241

## RESULT 13

D98032

type II site-specific deoxyribonuclease (EC 3.1.21.4) spnR-truncation [imported] - *Streptococcus pneumoniae*C:Species: *Streptococcus pneumoniae*

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: D98032

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnen, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellid, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Taskunas, S.R.

A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.

A:Reference number: A97872; M01D:21429245; PMID:11544234

A:Accession: D98032

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-60 &lt;KUR&gt;

A:Cross-references: GB:AE007317; PIDN:AL00089.1; PID:g15458927; GSPDB:GN00174

C:Genetics:

A:Gene: spnR-truncation

Query Match

Best Local Similarity 50.0%; Score 39.5; DB 2; Length 60;

Best Local Similarity 45.0%; Pred. No. 2.8;

Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

OY 2 EGSTDYVONI-----IOVAG 16

DB 21 EGENIDLYONKGRFLYKTAG 40

## RESULT 14

A35122

metalloproteinase (EC 3.4.-.-) mpr precursor, extracellular - *Bacillus subtilis*C:Species: *Bacillus subtilis*

C:Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 15-Oct-1999

C:Accession: A35122; I40010; A69660

R:Slama, A.; Rudolph, C.F.; Rufe Jr., G.A.; Sullivan, B.J.; Thieriault, K.A.; Alty, D.; F

J. Bacteriol. 172, 1024-1029, 1990

A:Title: Gene encoding a novel extracellular metalloproteinase in *Bacillus subtilis*.

A:Reference number: A35122; M01D:90130256; PMID:2105231

A:Accession: A35122

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-313 &lt;SLD&gt;

A:Cross-references: GB:L10505; NID:g143209; PIDN:AAA22604.1; PID:g143210; GB:M2903

R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.

Gene 70, 351-361, 1988

A:Title: Characterization of signal-sequence-coding regions selected from the Bacil

A:Reference number: 19994; M01D:89108019; PMID:3145906

A:Accession: I40010

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-60, 65, 'L', 67, 'S', 69, 'AQA' &lt;RES&gt;

A:Cross-references: GB:M2916; NID:g143701; PIDN:AAA22832.1; PID:g143702

R:Rust, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Allioni, G.; Azevedo, V.; J

C:Bron, S.; Brouillette, S.; Busch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M

A:Enllich, S.D.; Emerson, P.T.; Enlian, K.D.; Errington, J.; Fabret, C.; Ferrar

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galazzi, A.;

leach, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulo

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;

Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port

Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Sc

A:Authors: Schleich, S.; Schroeder, R.; Seifone, F.; Sekiguchi, J.; Sekowska, A.;

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yosh

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Dancho, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subt*

A:Reference number: A69580; M01D:98044033; PMID:9384377

A:Accession: A69660

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-313 &lt;KUN&gt;

A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CA012018.1; PID:e11

A:Experimental source: strain 168

C:Genetics:

C:Keywords: hydrolase

Query Match

Best Local Similarity 49.4%; Score 39; DB 2; Length 313;

Best Local Similarity 70.0%; Pred. No. 23;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 TXDYVONI 14

DB 300 TNDYFNNOY 309

## RESULT 15

S22127

protein kinase polo (EC 2.7.1.-) - fruit fly (*Drosophila melanogaster*)C:Species: *Drosophila melanogaster*

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 24-Sep-1999

C:Accession: S22127

R:Sunel, C.E.

submitted to the EMBL Data Library, November 1991

A:Reference number: S22127

A:Accession: S22127

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-576 &lt;SUN&gt;

A:Cross-references: EMBL:X63361; NID:g8355; PIDN:CAA44963.1; PID:g8356

C:Genetics:

A:Gene: Flybase:polo

A:Cross-references: Flybase:FBgn0003124

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase

C:Keywords: Atp; phosphotransferase

F:23-27/Domain: protein kinase homology &lt;KIN&gt;

Query Match

Best Local Similarity 49.4%; Score 39; DB 2; Length 576;

Best Local Similarity 53.8%; Pred. No. 47;

Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

us-09-737-297-3.rpr

```
QY      3  GSTXDVYQNIQYA 15
          | : | : | | : | |
Db      551 GVSXKDLVQKIRYA 563
```

Search completed: May 13, 2003, 15:58:34  
Job time : 48 secs

GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2003, 15:52:14 ; Search time 25 Seconds

(without alignments)  
26.545 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79

Sequence: 1 AEGSTXDYONIQYAG 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.5	56.3	455	NCAP_CVMJH	P03417 murine coro
2	42	53.2	387	GLGC_BACST	O08326 bacillus st
3	42	53.2	512	YBS6_YEAST	P39000 saccharomyc
4	41	51.9	169	YVJO_ECOLI	P77482 escherichia
5	41	51.9	240	XYNC_STRLI	P62220 streptomyce
6	39	49.4	313	MPR_BACST	P39790 bacillus su
7	39	49.4	576	POLO_DROME	P52304 drosophila
8	39	49.4	757	PPR2_MOUSE	O35398 mus musculu
9	38	48.1	149	YVIL_STAVU	P41368 staphylococ
10	38	48.1	250	GLGC_BACCL	P30522 bacillus ca
11	38	48.1	505	C4DB_DROME	O94579 drosophila
12	38	48.1	753	PEP2_HUMAN	O14830 homo sapien
13	38	48.1	1051	UBAL_WHEAT	P20973 triticum ae
14	38	48.1	1051	UBA2_WHEAT	P31251 triticum ae
15	38	48.1	1146	ITR1_ROME	O24247 drosophila
16	37.5	47.5	455	NCAP_CVM1	P45705 bacillus st
17	37	46.8	210	XYNA_BACST	O82947 chromatium
18	37	46.8	383	CYCR_CHRYI	P39126 bacillus su
19	37	46.8	423	IDH_BACST	P50808 human papil
20	37	46.8	604	VEL_HPV36	P09922 mus musculu
21	37	46.8	631	MXI_MOUSE	P23325 escheichia
22	37	46.8	728	ARPA_ECOLI	P54936 caenorhabdi
23	37	46.8	961	LIN2_CAEEL	P03416 murine coro
24	36.5	46.2	454	NCAL_CVM5	P18448 murine coro
25	36.5	46.2	454	NCAL_CVM5	P18448 murine coro
26	36.5	46.2	454	NCAP_CVM3	P18449 murine coro
27	36.5	46.2	454	NCAP_CVM5	P26278 roseobacter
28	36	45.6	352	CYCR_CVMS	P22897 homo sapien
29	36	45.6	1456	MANR_HUMAN	O08372 plasmodium
30	36	45.6	3135	S230_PLAFO	O28823 bacillus su
31	35	44.3	120	FOLB_BACST	P06504 bos taurus
32	35	44.3	177	CRBS_BOVIN	P22914 homo sapien
33	35	44.3	177	CRBS_HUMAN	P22914 homo sapien

34	35	44.3	213	1	XYNA_BACCI	P09850 bacillus ci
35	35	44.3	213	1	XYNA_BACSU	P18429 bacillus su
36	35	44.3	255	1	FLA5_PYRHO	O58286 pyrococcus
37	35	44.3	324	1	Y365_BUCAI	P57446 buchnera ap
38	35	44.3	380	1	GLGC_BACSU	P39122 bacillus su
39	35	44.3	392	1	CEMA_NEPOL	O91k22 nephrolepti
40	35	44.3	420	1	GLGC_AGRTU	P39669 agrobacteri
41	35	44.3	450	1	UDPE_GVLO	O98166 lacanobia o
42	35	44.3	478	1	YSP3_YEAST	P25036 saccharomyc
43	35	44.3	491	1	TRPE_NEIMA	O94x20 neisseria m
44	35	44.3	491	1	TRPE_NEIMB	P56995 neisseria m
45	35	44.3	491	1	TRPE_NEIMC	O94358 neisseria m

## ALIGNMENTS

## RESULT 1

ID NCAP\_CVMJH STANDARD; PRT; 455 AA.

AC P03417; 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Nucleocapsid protein.

OS Murine coronavirus MHV (strain JHM).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.

OX NCBI\_TaxID=11144;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=83272950; PubMed=6308569;

RT Skinner M.A., Siddell S.G.;

RT "Coronavirus JHM: nucleotide sequence of the mRNA that encodes

RT nucleocapsid protein."

RL Nucleic Acids Res. 11:5045-5054(1983).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC or send an email to license@lsb.ch).

CC EMBL; X00990; CAA25497.1; .

DR PIR; A04024; VHIHMJ.

DR InterPro: IPR001218; Corona\_nucleocap.

DR Pfam: PF00937; Corona\_nucleoca; 1.

DR Nucleocapsid.

OS Nucleocapsid.

Query Match 56.3%; Score 44.5; DB 1; Length 455;

Best Local Similarity 56.2%; Pred. No. 1.2;

Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

QY 1 AEGSTXDYONIQYAG 16

Db 340 ADGPTDYVE-LQYSG 354

GLGC\_BACST STANDARD; PRT; 387 AA.

ID O08326;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Glucose-1-phosphate adenyltransferase (EC 2.7.7.27) (ADP-glucose

DE synthase) (ADP-glucose pyrophosphorylase).

GN GLGC.

OS Bacillus stearothermophilus.

```

OC Bacteria; Firmicutes; Bacillales; Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FB814;
RA MEDLINE=97386405; PubMed=9244254;
RA Takata H., Takaha T., Okada S., Takagi M., Tanaka T.;
RT "Characterization of a gene cluster for glycogen biosynthesis and a
RT heterotrimeric ADP-glucose cyrophosphorylase from Bacillus
RT stercorophilus";
RU J. Bacteriol. 179:4689-4698(1997).
CC -1- CARBAMYLIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diposphate + ADP-glucose.
CC -1- PATHWAY: Glycogen biosynthesis, first step.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D87026; BAA19589.1;
DR InterPro: IPR001825; NTP_transferase.
DR Pfam: PF00483; NTP_transferase; 1.
DR PROSITE: PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE: PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE: PS00810; ADP_GLC_PYROPHOSPH_3; 1.
KW Glycogen biosynthesis; Transferase; Nucleotidyltransferase.
SQ SEQUENCE 387 AA; 43266 MW; 3660DB962442978 CRC64;

Query Match
Best Local Similarity 53.2%; Score 42; DB 1; Length 387;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EGSXTDYVYONIOY 14
DB 100 EGTANAVYONINY 112
||:|||||
|:|:|:|:|

```

```

RESULT 3
YEST_YEAST STANDARD; PRT: 512 AA.
ID YEST_YEAST
AC P39000;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 56.6 kDa protein in RAD51-UPB9 intergenic region.
GN YER096W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkeler-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sent P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yellon M., Botstein D., Davis R.W.;
RU Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 249-454 FROM N.A.
RC MEDLINE=92257587; PubMed=1581961;
RA Shinohara A., Ogawa H., Ogawa T.;
RT "Rad51 protein involved in repair and recombination in S. cerevisiae
RT is a RecA-like protein.";

```

```

Cell 69:457-470(1992).
CC -1- SIMILARITY: SOME, TO YEAST SKTS AND TO S.POMBE SPAC24B1.10C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18839; AAB64651.1;
DR EMBL: D10023; BAA20966.1; ALT_SEQ.
DR SGD: S0000898; SHC1.
KW Hypothetical protein.
FT CONFLICT 311 311 K -> R (IN REF. 2).
SQ SEQUENCE 512 AA; 56578 MW; 654D742482616658 CRC64;

Query Match
Best Local Similarity 53.2%; Score 42; DB 1; Length 512;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TADYVYONIOYA 15
DB 255 TODIYSNFOYA 265
||:|||||
|:|:|:|:|

```

```

RESULT 4
YAOQ_ECOLI STANDARD; PRT: 169 AA.
ID YAOQ_ECOLI
AC P77482;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yaoQ.
GN YAOQ OR B0426 OR Z0529 OR ECS0480.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562; 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregory J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RU Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federle N., Hyman R., Kalman S., Komp C., Kurai O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RU Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RA MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller D.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RU Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RA MEDLINE=21156231; PubMed=11258796;

```

```

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsuo E., Nakayama K., Murata T., Tanaka M., Ito T.,
RA Tida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.,
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
RN [5]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE-99420866; PubMed-10493123.
RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.:
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxyapatite chromatography."
RL Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: STRONG, TO H. INFLUENZAE H1034.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE000149; AAC73529.1; -
CC DR EMBL: U82664; AAB40182.1; -
CC DR EMBL: AE005222; AAG54776.1; -
CC DR EMBL: AF002551; BAB3903.1; -
CC DR Ecogene: EG13613; ya30.
CC Complete proteome.
KW SEQUENCE 169 AA; 19047 MW; 202279CC84D155C CRC64;
SQ
Query Match 51.9%; Score 41; DB 1; Length 169;
Best Local Similarity 53.3%; Pred. No. 1.7;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
OY 2 EGSTXDYVQNIQYAG 16
DB 81 EGSSLDVPEIVHSG 95
RESULT 5
XNMC STRLI STANDARD; PRT; 240 AA.
AC P26320:
ID XNMC STRLI STANDARD; PRT; 240 AA.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Endo-1,4-beta-xylanase C precursor (EC 3.2.1.8) (Xylanase C)
DE (1,4-beta-D-xylan xylanohydrolase C).
GN XLNC.
OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1916;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-80.
RC STRAIN-66 / 1326;
RX MEDLINE-92077439; PubMed-1743521;
RA Shreck F., Roy C., Yasuchi M., Morosoli R., Kluepfel D.:
RT "Sequences of three genes specifying xylanases in Streptomyces
RT lividans."
RL Gene 107:75-82(1991).
CC -1- FUNCTION: CONTRIBUTES TO HYDROLYSE HEMICELLULOSE, THE MAJOR
CC COMPONENT OF PLANT CELL-WALLS.
CC -1- CATALYTIC ACTIVITY: Endohydrolase of 1,4-beta-D-xylosidic
CC linkages in xylans.
CC -1- PATHWAY: xylan degradation.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY G (FAMILY 11 OF GLYCOSYL
CC HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: M64553; AAA26836.1; -
CC DR EMBL: A25307; CAA01768.1; -
CC DR PIR: J50591; J50591.
CC HSR: P09850; 1XNB.
CC DR InterPro: IPR001137; GH_11.
CC Pfam: PF00457; Glyco_Hydro_11; 1.
CC DR PRINTS: PR00911; GLHYDRLASE11.
CC DR PROSITE: PS00776; GLYCOSYL-HYDROL_F11_1; 1.
CC DR PROSITE: PS00777; GLYCOSYL-HYDROL_F11_2; 1.
CC KW Xylan degradation; Hydrolase; Glycosidase; Signal.
CC FT SIGNAL 1 49
CC FT CHAIN 50 240 ENDO-1,4-BETA-XYLANASE C.
CC FT ACT_SITE 134 134 NUCLEOPHILE (BY SIMILARITY).
CC FT ACT_SITE 226 226 PROTON DONOR (BY SIMILARITY).
CC SQ SEQUENCE 240 AA; 25673 MW; FC65415780142CA CRC64;
Query Match 51.9%; Score 41; DB 1; Length 240;
Best Local Similarity 42.9%; Pred. No. 2.5;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
OY 1 AEGSTXDYVQNIQY 14
DB 156 SDGTVDYDQITRY 169
RESULT 6
MPR_BACSU STANDARD; PRT; 313 AA.
AC P39790:
ID MPR_BACSU STANDARD; PRT; 313 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Extracellular metalloprotease precursor (EC 3.4.21.-).
DE MPR.
GN Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-6P241;
RX MEDLINE-90130256; PubMed-2105291;
RA Sloma A., Rudolph C.E., Ruto G.A. Jr., Sullivan B.J., Theriault K.A.,
RA Ally D., Pero J.:
RT "Gene encoding a novel extracellular metalloprotease in Bacillus
RT subtilis."
RL J. Bacteriol. 172:1024-1029(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RA Haga K., Ito H., Yasumoto K., Takahashi H., Yoshikawa H.:
RT "Sequence analysis of the 70kb region between 17 and 23 degree of the
RT Bacillus subtilis chromosome."
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE-98044033; PubMed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borzhet S.,
RA Borriest R., Bourcier L., Brans A., Braun M., Bridgell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Dentzot F., Devigne K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galiszi A., Galleron N.,
RA Gim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,

```

RA Guisepi G., Guy B.J., Haga K., Hafeez J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Kleer-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,  
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlon E., Scheller S., Schroeder R., Scottone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccout E., Takagi T., Takahashi H., Takemaru K.,  
 RA Tsuchiuchi M., Tamakoshi A., Tanaka T., Terpstra P., Vassaret A.,  
 RA Tostato V., Uchiyama S., Vandepol M., Vannier F., Vassaret A.,  
 RA Varti A., Wandt R., Wedler E., Wedler H., Weitzenger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.,  
 RA "the complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RA subtilis." ;  
 RA Nature 390:249-256(1997).  
 RL [4]  
 RN SEQUENCE OF 1-68 FROM N.A.  
 RP MEDLINE-69108019; PubMed-3145906;  
 RX Smith H., de Jong A., Bron S., Venema G.;  
 RA "Characterization of signal sequence-coding regions selected from the  
 RT *Bacillus subtilis* chromosome." ;  
 RL Gene 70:351-361(1988).  
 RN [5]  
 RP SEQUENCE OF 1-10 FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE-96118702; PubMed-7496533;  
 RA Saxild H.H., Jacobsen J.H., Nygaard P.;  
 RT "Functional analysis of the *Bacillus subtilis* prt gene encoding  
 RL formate-dependent glycanidase ribonucleotide transformylase." ;  
 RL Microbiology 141:2211-2218(1995).  
 CC -1 SUBUNIT: MONOMER.  
 CC -1 SUBCELLULAR LOCATION: Secreted.  
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY 52B.  
 CC -1 CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2  
 CC FAMILY OF SERINE PROTEASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L10505; AAA22604.1; -  
 DR EMBL: AB006424; AAA33121.1; -  
 DR EMBL: 269105; CAB12018.1; -  
 DR EMBL: M22916; AAA22832.1; -  
 DR PIR: A35122; A35122.  
 DR MEROPS: S01.272; -  
 DR Subtilisin; BGI0690; mpt.  
 DR InterPro: IPR000126; Ser-proteas-V8.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00039; V8PROTASE.  
 DR SMART: SM00020; TRYP-SPE; 1.  
 DR PROSITE: PS00672; V8-HIS; 1.  
 DR PROSITE: PS00673; V8-SER; 1.  
 DR Hydrolyase; Serine protease; Signal; Zymogen; Complete proteome.  
 FT SIGNAL 1 34  
 FT PROPEP 35 93  
 FT CHAIN 94 313 EXTRACELLULAR METALLOPROTEASE.  
 FT ACT\_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 131 147 BY SIMILARITY.  
 FT CONFID 61 68 OVSAPYEG -> PLESTAOA (IN REF. 4).  
 SO SEQUENCE 313 AA: 33842 MW; D41788EBD652AE94 CRC64;

Query Match 49.4%; Score 39; DB 1; Length 313;  
 Best Local Similarity 70.0%; Pred. NO. 8;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 TXYDONTQY 14  
 DB 300 TNDVFNNTQY 309  
 RESULT 7  
 ID POLO\_DROME STANDARD; PRT: 576 AA.  
 AC P52304; Q9VWB2;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Protein kinase polo (EC 2.7.1.-).  
 GN POLO OR CG12306.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Canton-S;  
 RX MEDLINE-92084090; PubMed-1660828;  
 RA Llamazares S., Moreira A., Tavares A., Giridhar C., Spruce B.A.,  
 RA Gonzalez C., Kares R.E., Glover D.M., Sunkel C.E.;  
 RT "Polo encodes a protein kinase homolog required for mitosis in  
 RL *Drosophila*." ;  
 RL Genes Dev. 5:2153-2165(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Berkeley;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Franke C., Baskley E.M.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Bolshakov S.,  
 RA Beeson K.Y., Beus P.V., Berman B.P., Bhandal D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Broxstein P., Brotler P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dudbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Gloadex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibergan C.,  
 RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasro F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy E., Murphy L., Murthy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palzozlo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen T.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;





```

ID      YI11:STAU      STANDARD:      PRT:      149 AA.
AC      P41369;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein in ILIS 5' region (ORF B) (Fragment).
OS      Staphylococcus aureus.
CC      Bacteria; Firmicutes; Bacillales; Staphylococcus.
CX      NCBI_TaxID=1280;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      SRRATN-12870;
RX      MEDLINE=94346383; PubMed=8067768;
RA      Hodgson J.E., Curnock S.P., Dyke K.G.H., Morris R.,
RA      Sylvester D.R., Gross M.S.;
RT      "Molecular characterization of the gene encoding high-level mupirocin
RT      resistance in Staphylococcus aureus J2870."
RL      Antimicrob. Agents Chemother. 38:1205-1208(1994).
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC
DR      EMBL; X75439; CAA5190.1; -
DR      InterPro; IPR001601; Methyltransf.
KW      Hypothetical protein.
FT      NON_TER
FT      SEQUENCE 149 AA; 17660 MW; 808FBEF150DBEC558 CRC64;
SQ
Query Match 48.1%; Score 38; DB 1; Length 149;
Best local similarity 58.3%; Pred. No. 5.3;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 4 STDVYQNIQYA 15
DB 83 STDVYKKNEXS 94
II IIII:II:
RESULT 10
GIGC_BACCL
ID GIGC_BACCL STANDARD; PRT: 250 AA.
AC P30522;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase) (Fragment).
GN GIGC.
OS Bacillus caldolyticus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1594;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93208370; PubMed=1296817;
RA Kiel J.A.K.W., Boels J.M., Beldman G., Venema G.;
RT "The g1gb gene from the thermophile Bacillus caldolyticus encodes a
RT thermostabile branching enzyme.";
RL DNA Seq. 3:221-232(1992)
CC -1- CARBLYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diphosphate + ADP-glucose.
CC -1- PATHWAY: Glycogen biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboratio
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no wa
CC modified and this statement is not removed. Usage by and for commercia

```

```

CC  entities requires a license agreement (See http://www.isb-sib.ch/announce
CC  or send an email to license@sib-sib.ch).
CC  -----
DR  EMBL: Z14057; CAA78441.1;
DR  PIR: S23857; S23857.
DR  InterPro: IPR001825; NTP_transferase.
DR  Pfam: PF00483; NTP_transferase: 1.
DR  PROSITE: PS00809; ADP_GLC_PYROPHOSPH. 1; 1.
DR  PROSITE: PS00810; ADP_GLC_PYROPHOSPH. 3; 1.
KW  Glycogen biosynthesis; transferase; Nucleosidyltransferase.
FT  NON_TER 250
SQ  SEQUENCE 250 AA; 28215 MW; A5AD3A087E3873F7 CRC64;

QY  2 EGSYADYQNYQY 14
    11: 111: 1
Db  100 EGSYADYQNY 112

Query Match 48.1%; Score 38; DB 1; Length 250;
Best local Similarity 46.2%; Pred. No. 9.5;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

RESULT 11
C4D8_DROME STANDARD; PRT; 505 AA.
AC Q9V573; Q24127;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 4d8 (EC 1.14.-.-) (CYP11D8).
GN CYP4D8 OR C4321.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Mandibulata, Pancrustacea, Hexapoda;
OC Insecta; Pterygota, Neoptera; Endopterygota, Diptera; Brachycera;
OC Muscomorpha; Ephyridroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;

[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Erlandson R.C., Rogers Y., H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A.A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhargava P., Brottier P.,
RA Borovica D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes W., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Galbraith W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heintz T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Krill C., Krautz S., Kulp D., Lai Z.,
RA Liou X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclow J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

```



Best Local Similarity 50.0%; Pred. No. 34;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 5 TXYDYONIQYAG 16  
| : : : : :  
Db 44 TWSTFQSTFAG 55

## RESULT 13

UBA1 WHEAT STANDARD; PRT; 1051 AA.  
ID UBA1 WHEAT  
AC P20973.  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ubiquitin-activating enzyme E1 1.  
GN UBA1.  
OS Triticum aestivum (wheat).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Triticum.  
OX NCBI\_TaxID=4565;  
RN (1)  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
RC STRAIN=cv. Augusta;  
RX MEDLINE=90368797; PubMed=2203788;  
RA Hatfield P.M., Vierstra R.D.;  
RT Cloning of ubiquitin activating enzyme from wheat and expression of  
RT a functional protein in Escherichia coli.;  
RL J. Biol. Chem. 265:15813-15817(1990).  
RN (2)  
RP MUTAGENESIS OF CYSTEINE RESIDUES AND ACTIVE SITE.  
RX MEDLINE=92340519; PubMed=1634524;  
RA Hatfield P.M., Vierstra R.D.;  
RT Multiple forms of ubiquitin-activating enzyme E1 from wheat.  
RT Identification of an essential cysteine by in vitro mutagenesis.;  
RL J. Biol. Chem. 267:14799-14803(1992).  
CC -1- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLATING WITH ATP ITS  
CC CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS  
CC RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING  
CC AN UBIQUITIN-E1 THIOLESTER AND FREE AMP.  
CC -1- PATHWAY: Ubiquitin conjugation; first step.  
CC -1- SUBUNIT: MONOMER.  
CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
CC -1- MISCELLANEOUS: THERE ARE TWO ACTIVE SITES WITHIN THE E1 MOLECULE,  
CC ALLOWING IT TO ACCOMMODATE TWO UBIQUITIN MOIETIES AT A TIME, WITH A  
CC NEW UBIQUITIN FORMING AN ADENYLATE INTERMEDIATE AS THE PREVIOUS  
CC ONE IS TRANSFERRED TO THE THIOLE SITE.  
CC -1- MISCELLANEOUS: THERE ARE MULTIPLE GENES ENCODING E1 IN WHEAT.  
CC -1- SIMILARITY: TO E1 IN OTHER SPECIES.  
CC -1- SIMILARITY: THE N-TERMINAL SHOWS SIMILARITY TO E. COLI CHLN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M55604; AAA34308.1; .  
CC DR PIR: A38373; A38373.  
CC DR PIR: A42873; A42873.  
CC DR InterPro: IPR000594; Ubact\_repeat.  
CC DR InterPro: IPR000127; Ubact\_repeat.  
CC DR InterPro: IPR000011; Uqin-activ enz.  
CC DR Pfam: PF00899; Thif. 2.  
CC DR Pfam: PF02134; Ubact. 1.  
CC DR PROSITE: PS00536; UBIQUITIN\_ACTIVAT.1; 1.  
CC DR PROSITE: PS00865; UBIQUITIN\_ACTIVAT.2; 1.  
CC UBL conjugation pathway; Ligase; Multigene family; Repeat.  
FT ACT SITE 626 626  
FT DOMAIN 56 605 2 APPROXIMATE REPEATS.

FT REPEAT 56 194 1-1.  
FT REPEAT 453 605 1-2.  
SQ SEQUENCE 1051 AA; 117007 MW; 9BA85FE3BACA621 CRC64;

Query Match 48.1%; Score 38; DB 1; Length 1051;  
Best Local Similarity 61.5%; Pred. No. 50;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0.

QY 3 GSTDYONIQYA 15  
| | | : : :  
Db 347 GSTDYONIQYEA 359

## RESULT 14

UBA2 WHEAT STANDARD; PRT; 1051 AA.  
ID UBA2 WHEAT  
AC P31251.  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ubiquitin-activating enzyme E1 2.  
GN UBA2.  
OS Triticum aestivum (wheat).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
OC Triticeae; Triticum.  
OX NCBI\_TaxID=4565;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92340519; PubMed=1634524;  
RA Hatfield P.M., Vierstra R.D.;  
RT Multiple forms of ubiquitin-activating enzyme E1 from wheat.  
RT Identification of an essential cysteine by in vitro mutagenesis.;  
RL J. Biol. Chem. 267:14799-14803(1992).  
CC -1- FUNCTION: ACTIVATES UBIQUITIN BY FIRST ADENYLATING WITH ATP ITS  
CC CARBOXY-TERMINAL GLYCINE RESIDUE AND THEREAFTER LINKING THIS  
CC RESIDUE TO THE SIDE CHAIN OF A CYSTEINE RESIDUE IN E1, YIELDING  
CC AN UBIQUITIN-E1 THIOLESTER AND FREE AMP.  
CC -1- PATHWAY: Ubiquitin conjugation; first step.  
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).  
CC -1- MISCELLANEOUS: THERE ARE TWO ACTIVE SITES WITHIN THE E1 MOLECULE,  
CC ALLOWING IT TO ACCOMMODATE TWO UBIQUITIN MOIETIES AT A TIME, WITH A  
CC NEW UBIQUITIN FORMING AN ADENYLATE INTERMEDIATE AS THE PREVIOUS  
CC ONE IS TRANSFERRED TO THE THIOLE SITE.  
CC -1- MISCELLANEOUS: THERE ARE MULTIPLE GENES ENCODING E1 IN WHEAT.  
CC -1- SIMILARITY: TO E1 IN OTHER SPECIES.  
CC -1- SIMILARITY: THE N-TERMINAL SHOWS SIMILARITY TO E. COLI CHLN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M50663; AAA34265.1; .  
CC DR InterPro: IPR000594; Thif\_domain.  
CC DR InterPro: IPR000127; Ubact\_repeat.  
CC DR InterPro: IPR000011; Uqin-activ enz.  
CC DR Pfam: PF00899; Thif. 2.  
CC DR Pfam: PF02134; Ubact. 1.  
CC DR PROSITE: PS00536; UBIQUITIN\_ACTIVAT.1; 1.  
CC DR PROSITE: PS00865; UBIQUITIN\_ACTIVAT.2; 1.  
CC UBL conjugation pathway; Ligase; Multigene family; Repeat.  
FT ACT\_SITE 626 626  
FT ACT\_SITE 626 626  
FT DOMAIN 56 605 2 APPROXIMATE REPEATS.  
FT REPEAT 56 194 1-1.  
FT REPEAT 453 605 1-2.  
SQ SEQUENCE 1051 AA; 116825 MW; B1F8370F6A299B73 CRC64;

Query Match 48.1%; Score 38; DB 1; Length 1051;  
Best Local Similarity 61.5%; Pred. No. 50;



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: May 13, 2003, 15:52:54 ; Search time 85 Seconds

(without alignments)  
38.785 Million cell updates/sec

Title: US-09-737-297-3

Sequence: 1 AEGSTXDYVNIQYAG 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.21.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	54.4	380	16 Q9CHN1	Q9CHN1 lactococcus
2	43	54.4	426	2 Q9I385	Q9I385 clostridium
3	42	53.2	380	16 Q97Q57	Q97Q57 streptococ
4	42	53.2	470	17 Q8T0M4	Q8T0M4 metanosarc
5	41	51.9	241	16 Q9RI72	Q9RI72 streptomyce
6	41	51.9	2659	5 Q9VB07	Q9VB07 diosopphila
7	40.5	51.3	209	2 Q68567	Q68567 xanthomonas
8	40.5	51.3	315	2 Q53010	Q53010 burkholderi
9	40.5	51.3	453	16 Q8R712	Q8R712 thermococ
10	40	50.6	116	3 Q9VY24	Q9VY24 gseuannomy
11	40	50.6	187	16 Q8XNMO	Q8XNMO clostridium
12	40	50.6	295	2 Q9ZHR4	Q9ZHR4 azospirillu
13	40	50.6	338	16 Q8RED1	Q8RED1 fusobacteri
14	40	50.6	455	16 Q9JR33	Q9JR33 neisseria m
15	40	50.6	458	3 Q9QW11	Q9QW11 pyromyces r
16	40	50.6	701	5 Q967D4	Q967D4 geodia cydo

17	40	50.6	724	5 Q21222	Q21222 caenorhabdi
18	40	50.6	16215	5 Q9NFS3	Q9NFS3 diosopphila
19	39	49.4	191	2 Q9EW89	Q9EW89 streptomyc
20	39	49.4	248	16 Q8XR03	Q8XR03 ralsconia s
21	39	49.4	301	5 Q8SRU2	Q8SRU2 encephalito
22	39	49.4	382	2 Q8RTW9	Q8RTW9 uncultured
23	39	49.4	623	16 Q8Y838	Q8Y838 listeria mo
24	39	49.4	683	13 Q9W6R4	Q9W6R4 fugu rubrip
25	39	49.4	711	12 Q9IC89	Q9IC89 human astir
26	39	49.4	711	12 Q9ICD1	Q9ICD1 human astir
27	39	49.4	711	12 Q9JH52	Q9JH52 human astir
28	39	49.4	711	12 Q9JH58	Q9JH58 human astir
29	39	49.4	711	12 Q9JH57	Q9JH57 human astir
30	39	49.4	711	12 Q9JH56	Q9JH56 human astir
31	39	49.4	711	12 Q9JH55	Q9JH55 human astir
32	39	49.4	711	12 Q9JH54	Q9JH54 human astir
33	39	49.4	711	12 Q9JH53	Q9JH53 human astir
34	39	49.4	711	12 Q9JH52	Q9JH52 human astir
35	39	49.4	711	12 Q67744	Q67744 human astir
36	39	49.4	829	5 P90754	P90754 caenorhabdi
37	39	49.4	842	12 Q91FX1	Q91FX1 human astir
38	39	48.7	545	3 Q12559	Q12559 aspergillus
39	38.5	48.7	796	4 Q9NR57	Q9NR57 homo sapien
40	38.5	48.7	796	4 Q9H096	Q9H096 homo sapien
41	38.5	48.7	796	4 Q960K1	Q960K1 homo sapien
42	38.5	48.7	796	4 Q9NKG2	Q9NKG2 homo sapien
43	38.5	48.7	796	4 Q9NKG2	Q9NKG2 homo sapien
44	38.5	48.7	796	11 Q9E0H3	Q9E0H3 mus musculu
45	38	48.1	151	5 Q8T4H9	Q8T4H9 diosopphila

## ALIGNMENTS

RESULT 1	Q9CHN1	PRELIMINARY:	PRT:	380 AA.
AC	Q9CHN1	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27).			
GN	Gluc OR L10697.			
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).			
OC	Bacteria: Firmicutes: Bacillu/Clostridium group; Lactobacillales.			
OX	NCBI_TaxID=1360;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-IL1403;			
RX	MEDLINE=2135186; PubMed=11337471;			
RA	Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,			
RA	Weisenbach J., Ehrlich S.D., Sorokin A.;			
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus			
RT	lactis ssp. lactis IL1403."			
RL	Genome Res. 11:731-753(2001).			
DR	EMBL: AEO06303; AAK04795.1; .			
DR	InterPro: IPR001451; Hexapep.transf.			
DR	InterPro: IPR001825; NTP-transferase.			
DR	Pfam: PF00132; hexapep. 2.			
DR	Pfam: PF00483; NTP-transferase: 1.			
DR	PROSITE: PS00808; ADP-GUC-PYROPHOSPH. 1.			
DR	PROSITE: PS00810; ADP-GUC-PYROPHOSPH. 3; UNKNOWN_1.			
KW	transferase; Nucleotidyltransferase; Complete proteome.			
SQ	SEQUENCE 380 AA; 42053 MW; 524E9934E59CF6F8 CRC64;			
QY	2 EGTSHAVYONISY 14	Score 43; DB 16; Length 380;		
Db	100 EGTSHAVYONISY 112	Best Local Similarity 61.5%; Pred. No. 20;		
		Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;		

## RESULT 2

ID	PRELIMINARY:	PRT:	426 AA.
AC	09L385;		
AD	01-OCR-2000 (TEMBirel, 15, Created)		
DT	01-OCR-2000 (TEMBirel, 15, Last sequence update)		
DT	01-MAR-2002 (TEMBirel, 20, Last annotation update)		
DE	Glucose-1-phosphate adenylyltransferase.		
GN	G1GC		
OS	Clostridium cellulolyticum.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia		
CC	Clostridiales; Clostridiaceae; Clostridium.		
CX	NCPI_TaxID=1521;		

	Query March	54.4%	Score 43	DB 2	Length 426	
	Best Local Similarity	61.5%	Pred. No. 23			
	Matches	8	Conservative	2	Mismatches	3
					Indels	0
					Gaps	0
QY	2 EGSTKDPVXNNIY 14					
	::: :::::					
DB	101 KGTANAVYNNIY 113					

## RESULT

```

AC 097057; PRELIMINARY; PRT; 380 AA.
AD 097057;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase.
GN Sp1122.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
CX NCBI_TaxId=13313.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA MEDLINE=11463916;
RA Rettelinn H., Nelson K.E., Pallesen I.T., Eisen J.A., Read I.D.,
RA Petersen S., Heideberg J., Debey R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolony J.F., Nelson W.C., Peterson J.D.,
RA Umeyan L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiolini S., Dickinson T., Hickey E.K.,
RA Holt I.E., Lotius B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL: AE007413; AAK75233.1; .
DR TIGR: Sp1122; -.
DR InterPro: IPR001825; NTP_transferase.
DR Pfam: PF00483; NTP_transferase; 1.

```

## RESULT 4

```

Query Match          53.2%; Score 42; DB 16; Length 380;
Best Local Similarity 53.8%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0

CY      2  EGSTXDYQNIQY 14
      |||:::|||||
Db      100  EGSTHAIVQNIQY 112

```

## AC Q8T

```

DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Transport protein.
CN MA1516.
OS Methanosarcina acetivorans.
OC Archaea: Euryarchaeota: Methanococci: Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760. PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atwood D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArlano K., Johnson R.,
RA Linton L., McLean P., McKernan K., Galamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Heddlelich R., Ingram-Smith C., Kretzner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Perry J.G., Jarrell K.F., Jing H., Macario A.-J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AB010821; AAA04930.1; -.
KW Complete proteome.
SQ SEQUENCE 470 AA: 50953 MW: 4CD45B2299C32B9D CRC64;

Query Match 53.2%; Score 42; DB 17; Length 470;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 GSTYDVTQNTQYAG 16
Db 390 GSTVSLYNSMKYAG 403
||| : ||::|||

RESULT 5
ID 09R172 PRELIMINARY: PRT: 241 AA.
AC 09R172;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Endo-1,4-beta-xylosidase.
GN XLNC OR SC00105 OR SCJ11.34C.
OS Streptomyces coelicolor.
OC Bacteria: Firmicutes: Actinobacteria: Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A5(2);

```

```

RA Oliver K., Harris D.;
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Godle A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabdinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL; AL109949; CAB52919.1; -
DR HSSP; P09850; 1XNB.
DR InterPro; IPR001137; GH_11.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR PRINTS; PR00911; GHYDRASBL.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
KM Glycosylase: Hydrolase; Xylan degradation.
SQ SEQUENCE 241 AA; 25845 MW; 37B6E99DE51B78F CRC64;

Query Match 51.9%; Score 41; DB 16; Length 241;
Best Local Similarity 42.9%; Pred. NO. 27;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYVONTIOY 14
Db 157 SDGTYDIYOTRY 170

RESULT 6
OYVBUT ID PRELIMINARY; PRT; 2659 AA.
AC OYVBUT;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG11856 protein.
CN CG11856
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Plekoyota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
CX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRKELEY.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chan L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,

```

```

RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.C.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Butler H., Cadieu E., Center A., Chandra I.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gladex A., Gong F., Gotrell J.H., Gu Z., Guan P., Harris M.,
RA Harlow N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jastli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Switskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Welschenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003731; AAF56430.1; -
DR HSSP; P49792; 1RRP.
DR Flybase; FBgn0039302; CG11856.
DR InterPro; IPR000697; RanBP1_NASP.
DR InterPro; IPR000156; RanBP1.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR001876; Znf_RangDP.
DR Pfam; PF00638; Ran_BP1_4.
DR Pfam; PF00641; Zf-RanBP_2.
DR SMART; SM00160; RanBP_4.
DR SMART; SM00547; Znf_RBZ_2.
DR PROSITE; PS01358; ZF_RANBP2_1; UNKNOWN.2.
SQ SEQUENCE 2659 AA; 292235 MW; BE406AD434D9BE0B CRC64;

Query Match 51.9%; Score 41; DB 5; Length 2659;
Best Local Similarity 53.3%; Pred. NO. 3.9e+02;
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 2 EGSTXDYVONTIOYAG 16
Db 484 EGFGQDVYKQNLNYCG 498

RESULT 7
O68567 ID PRELIMINARY; PRT; 209 AA.
AC O68567;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE Restriction endonuclease R. XbaI.
CN XBAIR.
OS Xanthomonas campestris.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
CX NCBI_TaxID=339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BADRII;
RA Zhang B.-H., Wilson G.G.;

```



RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF051092; AAC08983.1; -  
 DR REBASE: 2126; Xbat.  
 KW Endonuclease.  
 SO SEQUENCE 209 AA; 23834 MW; 6D3CB8CE3FA92061 CRC64;

Query Match 51.3%; Score 40.5; DB 2; Length 209;  
 Best Local Similarity 50.0%; Pred. No. 29;  
 Matches 10; Conservative 1; Mismatches 4; Indels 5; Gaps 1;

OY 2 EGSTXDYVYON-----IOYAG 16  
 || ||||| :|||  
 DB 49 EGRIDYVYQNGRFLYKYAG 68

RESULT 8  
 ID 053010 PRELIMINARY; PRT; 315 AA.

AC 053010;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Hydroxyquinol-1, 2-dioxygenase.  
 GN HADC.  
 OS Burkholderia pickettii (Pseudomonas pickettii).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 CC Ralstonia.  
 CX NCBI\_TaxID=329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DTP0602;  
 RA Hatta T., Nakano O., Takizawa N., Kiyohara H.,  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-27 FROM N.A.  
 RC STRAIN=DTP0602;  
 RA Takizawa N., Yokoyama H., Yanagihara K., Hatta T., Kiyohara H.,  
 RT "A locus of Pseudomonas pickettii DTP0602, had, that encodes 2,4,6-  
 trichlorophenol-4-dechlorinase with hydroxylase activity, and  
 RT hydroxylation of various chlorophenols by the enzyme.";  
 RL J. Ferment. Bioeng. 80:318-326(1995).  
 DR EMBL: D86544; BAA13107.1; -  
 DR HSSP; P00437; 3PCC.  
 DR InterPro; IPR000627; Dioxygenase.  
 DR Pfam; PF00775; Dioxygenase; 1.  
 KW Dioxygenase.  
 SO SEQUENCE 315 AA; 34592 MW; B3ED3EF470406CA0 CRC64;

Query Match 51.3%; Score 40.5; DB 2; Length 315;  
 Best Local Similarity 42.9%; Pred. No. 45;  
 Matches 9; Conservative 4; Mismatches 1; Indels 7; Gaps 1;

OY 3 GSTXDYVY-----NIOYAG 16  
 :|| |||: :|||  
 DB 143 GATIDVWQADADGRYDYAG 163

RESULT 9  
 ID 08R712 PRELIMINARY; PRT; 453 AA.

AC 08R712;  
 DT 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Periplasmic protease.  
 GN PRC3 OR TIE2617.  
 OS Thermoanaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.  
 CX NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MBT / JCM1007;

RX MEDLINE-21992816; PubMed=11997335;  
 RA Bao Q., Tian Y., Li W., Xu Z., Xian Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome.";  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL: AE013202; AAM25737.1; -  
 KW Protease; Complete proteome.  
 SO SEQUENCE 453 AA; 49953 MW; 106503DE3FE58A74 CRC64;

Query Match 51.3%; Score 40.5; DB 16; Length 453;  
 Best Local Similarity 64.3%; Pred. No. 68;  
 Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

OY 2 EGSTXDYVY-ONIOY 14  
 || ||||| :|||  
 DB 345 EGLTDPYVYKNVY 358

RESULT 10  
 ID 09UVZ4 PRELIMINARY; PRT; 116 AA.

AC 09UVZ4;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
 DE Xylanase (Fragment).  
 GN AXVL2.  
 OS Gaeananomyces graminis.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariomycetes incertae sedis; Magnaporthaceae; Gaeananomyces.  
 CX NCBI\_TaxID=29850;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Emami K.;  
 RT "PCR-based characterization of fungal xylanase genes.";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ249160; CAB53513.1; -  
 DR HSSP; P09850; 1XNB.  
 DR InterPro; IPR001137; GH\_11.  
 DR Pfam; PF00457; Glyco\_hydro\_11; 1.  
 DR PRINTS; PR00911; GLHYDRLASE1.  
 DR PROSITE; PS00776; GLYCOSYL\_HYDROL\_F11\_1; 1.  
 FT NOW\_TER 1  
 FT NON\_TER 1  
 FT 116 116  
 SO SEQUENCE 116 AA; 12791 MW; DA7BA4FEA6770E9E CRC64;

Query Match 50.6%; Score 40; DB 3; Length 116;  
 Best Local Similarity 42.9%; Pred. No. 18;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 AEGSTXDYVYONIOY 14  
 :|| ||||| :|||  
 DB 58 SDGGTYDLYGSTRY 71

RESULT 11

ID 08XNWO PRELIMINARY; PRT; 187 AA.

AC 08XNWO;  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical protein CPE0222.  
 GN CPE0222.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 CX NCBI\_TaxID=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / TYPE A;  
 RX PubMed=11792842;

```

RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kihara S., Hayashi H.,
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater."
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL: AP003185; BAB79928.1;
DR TIGRFAMs: TIGR01076; sortase_fam.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 187 AA; 20996 MW; A0CCCB5B313B2C CRC64;

Query Match
Best Local Similarity 50.6%; Score 40; DB 16; Length 187;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 AEGSTXDYONIQY 11
DB 77 AEGSTLDVLEN 87

RESULT 12
O9ZHR4 PRELIMINARY; PRT; 295 AA.
AC O9ZHR4;
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Cytbb3 cytochrome c oxidase CytP subunit.
GN CytP.
OS Azospirillum brasilense.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Azospirillum.
OX NCBI_TaxID=192;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP7;
RX MEDLINE=99009004; PubMed=9791120;
RA Marchal K., Sun J., Keljers V., Haaker H., Vanderleyden J.;
RT "A cytochrome cbb3 (Cytochrome c) terminal oxidase in Azospirillum
RT brasilense SP7 supports microaerobic growth."
RL J. Bacteriol. 180:5689-5696(1998).
DR EMBL: AF054871; AAC72074.1;
DR HSSP; P57736; 1C60.
DR InterPro: IPR004678; CCoP.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR003088; Cyt_C1.
DR InterPro: IPR002329; Cyt_C1C.
DR Pfam: PF00034; cytochrome_c; 2.
DR PRINTS: PR00505; CYTOCHROME_C1C.
DR TIGRFAMs: TIGR00782; CCoP.1.
DR PROSITE: PS00190; CYTOCHROME_C; 2.
SQ SEQUENCE 295 AA; 31807 MW; 8F3BD57A6E25F0F CRC64;

Query Match
Best Local Similarity 50.6%; Score 40; DB 2; Length 295;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 GSTXDYONIQY 14
DB 147 GTTADIVKTION 158

RESULT 13
O8RED1 PRELIMINARY; PRT; 338 AA.
AC O8RED1;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein FN1177.
GN FN1177.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586.
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykdis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
RA Larsen N., D'Souza M., Malinas T., Pusch G., Haselkorn R.,
RA Forstein M., Kripides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010623; ALU95373.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 338 AA; 40819 MW; CA00AA9225E885A2 CRC64;

Query Match
Best Local Similarity 50.6%; Score 40; DB 16; Length 338;
Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 2 EGSTXDYONIQY 14
DB 118 EGASFNRYRLRY 130

RESULT 14
O9JR33 PRELIMINARY; PRT; 455 AA.
AC O9JR33;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hypothetical protein NMA1831.
GN NMA1831 OR RTH42.
OS Neisseria meningitidis (serogroup A), and
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699; 487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
RA Jajelski K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20187481; PubMed=10722605;
RA Klee S.R., Nassif X., Kusecek B., Merker P., Beretti J.L., Achtman M.,
RA Tinsley C.R.;
RT "Molecular and biological analysis of eight genetic islands that
RT distinguish neisseria meningitidis from the closely related pathogen
RT neisseria gonorrhoeae."
RL Infect. Immun. 68:2082-2095(2000).
DR EMBL: AL162757; CAB85056.1;
DR EMBL: AJ391256; CAB72000.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 455 AA; 49625 MW; 706E6F3F56238064 CRC64;

Query Match
Best Local Similarity 50.6%; Score 40; DB 16; Length 455;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 4 STXDYONIQYAG 16
DB 352 SADVYCRIEVAG 364

```

RESULT 15

```

090W11      PRELIMINARY;      PRT;      458 AA.
ID 090W11;
AC 090W11;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Exocellolohydrolase Cbh120.
CN Cbh120.
OS Piromyces rhiziniflatus.
OC Eukaryota; Fungi; Chytridiomycota; Neocallimastigales;
OC Neocallimastigaceae; Piromyces.
OX NCBI_TaxID=73428;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2301;
RA Liu J.-H., Qiu X., Cheng K.-J.;
RT "Cloning of cellulase genes from ruminal fungus Piromyces
RT rhiziniflatus."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF174361; AD51054.2; -.
DR HSSP: P07987; ICB2.
DR InterPro: IPR002883; CBD_5.
DR InterPro: IPR001524; GH_6.
DR Pfam: PF02013; CBM_10; 2.
DR Pfam: PF01341; Glyco_hydro_6; 1.
DR PRINTS: PR00733; GLHYDRLASE6.
DR ProDom: PD003733; GH_6; 1.
KW Hydrolase.
SQ SEQUENCE 458 AA; 50360 MW; BF8CFB4F3C3930BE CRC64;

Query Match          50.6%; Score 40; DB 3; Length 458;
Best Local Similarity 53.8%; Pred. No. 84;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 4 STXDYQNIQYAG 16
   ||::|||
Db 132 STGNITNSQYQG 144

```

Search completed: May 13, 2003, 15:57:43  
 Job time : 89 secs